Freie Universität Berlin

Fachbereich Wirtschaftswissenschaft

Human Capital Returns to Education

Three Essays on the Causal Effects of Schooling on Skills and Health

Inaugural Dissertation

vorgelegt von

Sarah Christina Dahmann, M.Sc.

geboren in Aachen

zur Erlangung des akademischen Grades

doctor rerum politicarum

(Doktor der Wirtschaftswissenschaft)

Berlin, 2016

Gedruckt mit Genehmigung des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

Dekan: Prof. Dr. Dr. Andreas Löffler
Erstgutachterin: Prof. Dr. C. Katharina Spieß
Zweitgutachterin: Prof. Dr. Silke Anger
Tag der Disputation: 14. Dezember 2016

Ko-Autorenschaften und Vorveröffentlichungen

Chapter 2

- Koautorin: Silke Anger (IAB Nürnberg, Universität Bamberg, IZA)
- Vorveröffentlichungen:
  - IZA Discussion Paper No. 8139
  - SOEPpapers 658
  - IAB-Discussion Paper 29/2014
  - CINCH working paper 2015/05

Chapter 3

- Kein Koautor
- Vorveröffentlichungen:
  - SOEPpapers 769
  - CINCH working paper 2015/04

Chapter 4

- Koautor: Daniel Schnitzlein (Leibniz Universität Hannover, DIW Berlin)
- Keine Vorveröffentlichungen

Abstract

This dissertation estimates returns to education in terms of skills and health as important aspects of human capital given their importance in determining economic outcomes. The dissertation is based on three independent empirical research articles which constitute the three main chapters 2, 3, and 4, and a comprehensive introduction in Chapter 1 and conclusion in Chapter 5. The analyses exploit several exogenous variations in German educational policies and apply state-of-the art microeconometric techniques for causal inference. The chapters complement each other in three dimensions: First, each chapter looks at different aspects of education, together considering both qualitative measures – including the curriculum, learning intensity, and timing – as well as quantitative factors – like years of schooling. Second, while all chapters estimate non-monetary effects of education, each investigates a specific dimension of human capital, overall addressing cognitive skills, non-cognitive skills, and health. Third, to complement each other from a life-course perspective, different age groups are analyzed including seventeen-year old adolescents still enrolled in secondary education, young adults at the time of graduating from high school, and adults between the ages 50 and 85, long after the completion of secondary education.

Chapter 2 investigates the short-term effects of a reduction in the length of high school on students' personality traits using a school reform carried out at the state level in Germany as a quasi-natural experiment. Starting in 2001, academic-track high school education was reduced by one year in most of Germany's federal states, with the overall curriculum left unchanged. This enabled students to obtain their university entrance qualification after only 12 rather than 13 years of schooling, but it increased learning intensity through an increase in weekly class hours. Exploiting the variation in the length of high school over time and across states, the effect of learning intensity on students' Big Five personality traits and on their locus of control is identified. Using representative data on adolescents and young adults from the German Socio-Economic Panel (SOEP) study, the estimates show that shortening high school caused students to be less emotionally stable. Furthermore, the results point to

important heterogeneous effects. In addition to regional and gender differences, students not living with both parents and students with migration background showed stronger personality changes following the reform: they became more open and extroverted, and more open and conscientious, respectively. Thus, the chapter concludes that the educational system plays an important role in shaping adolescents' personality traits.

Chapter 3 focuses on cognitive skills and investigates two mechanisms through which they may be affected by education. I rely on the same reform analyzed in Chapter 2 for causal identification but conduct two separate analyses based on different datasets: 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 SOEP data on seventeen year-olds. Second, I investigate the influence of earlier instruction at younger ages, using data from the German National Educational Panel Study (NEPS) 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.

Chapter 4 analyzes whether education has a protective effect on mental health. To estimate causal effects, an instrumental variable (IV) technique is implemented with two different instruments to estimate local average treatment effects (LATE) at different parts of the educational distribution: (i) a reform extending compulsory schooling by one year implemented in West German federal states between 1949 and 1969 and (ii) the individual availability of higher education measured by the spatial distance to the nearest university at age 19. The analyses are based on rich individual SOEP data on adults aged 50 to 85, augmented by detailed information on universities from the German Rectors' Conference. Analyses on the Mental Component Summary (MCS) score as a generic measure of overall mental health are complemented by disorder-specific diagnoses. Results support existing evidence on a positive relationship between completed years of secondary schooling and mental health in standard Ordinary Least Squares (OLS) estimations. However, the IV estimations reveal no such causal protective effect. If any, the estimates point towards a negative effect among the lower educated. These results are confirmed when explicitly modeling effect heterogeneity through marginal treatment effects.

Keywords: Education, Learning Intensity, Instructional Time, Timing of Instruction, Human Capital, Skill Formation, Personality Traits, Cognitive Skills, Mental Health, High School Reform, Compulsory Schooling Reform, Geographic Proximity to College, Difference-in-differences, Instrumental Variables, Marginal Treatment Effects

Zusammenfassung

Diese Dissertation untersucht den kausalen Einfluss von Bildung auf individuelle Fähigkeiten und Gesundheit. Diese beiden Bildungserträge stellen wichtige Elemente des Humankapitals dar und sind somit von herausragender Bedeutung in Bezug auf den Arbeitsmarkt, sowohl für den individuellen Erfolg als auch für die gesamtwirtschaftliche Analyse.

Die Dissertation setzt sich dabei aus drei empirischen Forschungsartikeln zusammen, welche in den drei Hauptkapiteln 2, 3, und 4 behandelt werden, sowie aus einer ausführlichen Einleitung in Kapitel 1 und einer Schlussfolgerung in Kapitel 5. Die Identifikation kausaler Effekte von Bildung auf die individuellen Fähigkeiten sowie die Gesundheit erfolgt mittels räumlicher und zeitlicher Variation infolge verschiedener Reformen in der deutschen Bildungspolitik. Zur Schätzung der Effekte werden aktuelle mikroökonometrische Methoden verwendet. Die einzelnen Kapitel ergänzen sich in dreifacher Hinsicht: Erstens beleuchtet jedes Kapitel unterschiedliche Aspekte von Bildung. Hierbei werden sowohl qualitative Aspekte, wie das Curriculum, die Lernintensität und die zeitliche Abstimmung des Lehrplans, als auch quantitative Aspekte, wie die Anzahl an Schuljahren, berücksichtigt. Zweitens analysiert jedes Kapitel eine separate Dimension des Humankapitals. Während alle Kapitel nicht-monetäre Bildungserträge untersuchen, liegt der spezifische Fokus der verschiedenen Kapitel jeweils auf kognitiven Fähigkeiten, nicht-kognitiven Fähigkeiten und mentaler Gesundheit. Drittens werden sowohl siebzehnjährige Jugendliche, die noch die Schule besuchen, junge Erwachsene im Abiturienten-Alter und Erwachsene im Alter zwischen 50 und 85 Jahren, deren Schulausbildung schon lange zurück liegt, betrachtet. Durch diese parallele Analyse verschiedener Altersgruppen sind verschiedene Phasen des Lebensverlaufs abgebildet.

Kapitel 2 analysiert kurzfristige Effekte einer Reduzierung der Gesamtschulzeit auf Persönlichkeitseigenschaften. Hierzu wird die sogenannte G8-Reform verwendet, die die Gymnasialzeit von neun auf acht Jahre verkürzte. Da die Reform in verschiedenen Bundesländern zu verschiedenen Zeitpunkten, beginnend im Jahr 2001, durchgeführt worden ist, stellt sie ein quasinatürliches Experiment dar. Durch die Verkürzung der Schulzeit bei gleichzeitiger Beibehaltung

des Umfangs des Unterrichtsstoffs stieg die Anzahl an Wochenstunden und somit die Lernintensität entsprechend an. Mithilfe der zeitlichen und räumlichen Variation in der Gesamtdauer des Gymnasiums kann der Effekt der Lernintensität auf die Persönlichkeitseigenschaften von SchülerInnen identifiziert werden. Die Persönlichkeitseigenschaften werden hierbei mithilfe des Fünf-Faktoren Modells und der internalen Kontrollüberzeugung erfasst. Die Schätzungen auf Basis repräsentativer Daten des Sozio-oekonomischen Panels (SOEP) zu Jugendlichen und jungen Erwachsenen zeigen, dass die gesteigerte Lernintensität eine Verringerung der emotionalen Stabilität der SchülerInnen bewirkt. Außerdem weisen die Ergebnisse wichtige Heterogenitäten auf. Neben regionalen und Geschlechterunterschieden, zeigen sich insbesondere bei SchülerInnen, die nicht bei beiden Eltern aufwachsen, und SchülerInnen mit Migrationshintergrund deutlichere Persönlichkeitsveränderungen infolge der erhöhten Lernintensität: für die erste Gruppe ein höheres Maß an Offenheit und Extrovertiertheit, für die zweite Gruppe ein höheres Maß an Offenheit und Gewissenhaftigkeit. Dieses Kapitel schlussfolgert daher, dass das Bildungssystem eine wichtige Rolle in der Persönlichkeitsbildung Jugendlicher spielt.

Kapitel 3 betrachtet zwei Mechanismen um zu untersuchen, wie Bildung kognitive Fähigkeiten beeinflusst. Um kausale Effekte zu identifizieren, beruht die Analyse auf derselben Reform, die bereits in Kapitel 2 als Identifikationsstrategie genutzt wurde. Insbesondere werden in Kapitel 3 zwei separate Analysen durchgeführt: Zum einen wird die Variation über die Zeit und zwischen Bundesländern ausgenutzt, um den Effekt eines Anstiegs in der Anzahl der Unterrichtsstunden zu beurteilen. Dies geschieht auf Basis von Daten des SOEP zu siebzehnjährigen Jugendlichen. Zum anderen wird der Einfluss des Zeitpunkts des Unterrichts im Lebenslauf der SchülerInnen betrachtet. Auf Basis von Daten des Nationalen Bildungspanels (NEPS) zu AbiturientInnen wird untersucht, inwiefern sich die Vermittlung desselben Lernstoffs in einem jüngeren Alter auf kognitive Kompetenzen auswirkt. Die Ergebnisse zeigen, dass Bildung vor allem durch die Anzahl an Unterrichtsstunden auf kognitive Fähigkeiten wirkt, und weniger durch eine zeitliche Verschiebung des Unterrichts im Hinblick auf das Alter der SchülerInnen. Es zeigt sich auch, dass ein Anstieg an Unterrichtszeit Geschlechterunterschiede, insbesondere im Bereich der numerischen Fähigkeiten, deutlich verschärft.

Kapitel 4 untersucht, ob sich Bildung protektiv auf mentale Gesundheit auswirkt. Um kausale Effekte zu identifizieren, wird eine Instrumentvariablen (IV) Schätzung angewandt. Lokale Durchschnittseffekte für verschiedene Bereiche der Bildungsverteilung werden mithilfe zweier unterschiedlicher Instrumentenvariablen geschätzt: Zum einen wird die Variation in der zeitlichen Ausweitung der Schulpflicht um ein Jahr, die zwischen 1949 und 1969 in westdeutschen Bundesländern zu verschiedenen Zeitpunkten durchgeführt wurde, herangezogen.

Zum anderen wird die Variation in der Verfügbarkeit höherer Bildungsmöglichkeiten anhand der räumlichen Distanz zur nächstgelegenen Universität oder Fachhochschule im Alter von 19 Jahren genutzt. Die Analysen beruhen auf reichhaltigen Daten des SOEP zu Erwachsenen im Alter zwischen 50 und 85 Jahren, ergänzt durch detaillierte Informationen zu Universitäten und Fachhochschulen der Stiftung zur Förderung der Hochschulrektorenkonferenz. Mentale Gesundheit wird durch den „Mental Component Summary (MCS) Score“ abgebildet, der allgemeine mentale Gesundheit auf einer kontinuierlichen Skala beschreibt, und durch Informationen zu medizinischen Diagnosen spezifischer psychischer Krankheiten ergänzt. Die Ergebnisse einer Kleinste-Quadrate-Schätzung bestätigen die bestehende Evidenz zur positiven Korrelation zwischen Bildung und mentaler Gesundheit. Die IV Schätzungen zeigen jedoch, dass diese nicht kausal zu interpretieren ist. Wenn überhaupt, deuten die Ergebnisse eher auf einen negativen Effekt hin, zumindest am unteren Rand der Bildungsverteilung. Dieser Befund wird in „Marginal Treatment Effect“-Analysen bestätigt, die explizit heterogene Effekte entlang der intrinsischen Präferenz für Bildung modellieren.

Schlüsselwörter: Bildung, Lernintensität, Unterrichtszeit, zeitliche Abstimmung des Unterrichts, Humankapital, Entstehung von Fähigkeiten, Persönlichkeitseigenschaften, Kognitive Fähigkeiten, Mentale Gesundheit, G8-Reform zur Schulzeitverkürzung, Ausweitung der Schulpflicht, Räumliche Distanz zur Universität, Differenzen-von-Differenzen Schätzer, Instrumentvariablen Schätzer, Marginal Treatment Effects

Acknowledgments

First and foremost, I am especially grateful to my two supervisors. I thank Prof. C. Katharina Spieß for accepting me as a doctoral student and perfectly handling the balance between guidance and freedom I needed to complete this dissertation. Further, she did not only give me valuable feedback whenever I needed assistance, but also integrated me in a very stimulating research environment. The monthly meetings with all of her doctoral students have continuously benefited the progress of my dissertation and being introduced through her to research networks focusing on education has greatly stimulated my interest in the topic. And I thank Prof. Silke Anger, who is not only my second supervisor, but has been a mentor for me throughout my entire time as a doctoral student. I am incredibly grateful for all very fruitful discussions, and especially her great personal support and guidance in my academic decision making.

Two chapters of this dissertation are based on joint work, for which I am very grateful to my co-authors Silke Anger and Daniel Schnitzlein. Working closely together with them, I learned a lot from their accuracy and critical thinking, but above all I very much enjoyed the collaborations and am thankful for all their scientific input and for sharing their knowledge with me.

I could not have completed this dissertation without the support of many other persons. I would like thank all my colleagues at DIW and in particular in the SOEP department for all fruitful discussions and valuable input. A special thanks goes to Adrian Hille and Daniel Kemptner for all the time you took for extensive and very helpful discussions, and to Adam Lederer and Deborah Anne Bowen for proof-reading parts of this dissertation. I thank Jürgen Schupp for welcoming me in the SOEP department where I enjoyed an interdisciplinary stimulating research environment and I particularly appreciated all data-related assistance. I am also thankful to the Graduate Center and Helmut Lütkepohl for providing an excellent structured PhD program I enjoyed being part of. I am very grateful to the German Academic Scholarship Foundation (*Studienstiftung des deutschen Volkes*) for awarding me a scholarship which enabled me to dedicate most of my time to working on this dissertation. The last chapter of this disser-

tation is part of the project *Nicht-monetäre Erträge von Bildung in den Bereichen Gesundheit, nicht-kognitive Fähigkeiten sowie gesellschaftliche und politiche Partizipation* which was funded by the German Federal Ministry of Education and Research. Further, generous funding by DIW enabled me to present my work at national and international conferences and workshops. I thank all seminar and conference participants, and guests at DIW, who commented on my work. I also thank my professors and mentors at Maastricht University, where I completed my Bachelor and Master studies, for excellently preparing me for this and awakening my interest in pursuing a PhD in the first place.

I thank my family for their confidence and encouragement. My grandparents', Trude and Alois, interest in my work has always boosted my motivation and helped my own understanding when trying to simplify explanations in non-economist's terms. And most importantly, I thank my parents, Inge and Peter, who have always supported me. I am deeply grateful for everything you made possible for me, for securing and accompanying my entire education, and for your unconditional support and trust in any decision that I took. Without you, I would not be who and where I am. For all that, I dedicate this dissertation to you.

Finally, I would like to thank all my colleagues at DIW who really made me enjoy my time as a doctoral student. The past years would not have been the same without you and I am grateful for all friends that I made. My office mates Doreen Triebe and Christian Krekel not only were incredibly valuable in discussing all bigger and smaller scientific issues, but especially the sharing and laughing together in the office made my days. And I thank Kristina Krell, Michael Neumann, Ulrich Schneider, Stefan Seifert, Rainer Siegers, Andreas Thiemann, Songül Tolan, Christian Zankiewicz, and Alexander Zerrahn for all the delicious lunches in the mensa, the coffee and chocolate breaks, the glorious table soccer matches and your patience in teaching me, and the afterwork-beers to not forget that there is a life outside academics.

Berlin, October 2016

Sarah Dahmann

Rechtliche Erklärung

Erklärung gem. § 4 Abs. 2 (Promotionsordnung)

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

Berlin, Oktober 2016
Sarah Dahmann

Erklärung gem. § 10 Abs. 3 (Promotionsordnung)

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe: Software LaTeX, Stata, R und QGIS, Literatur siehe Literaturverzeichnis. Auf dieser Grundlage habe ich die Arbeit selbstständig verfasst.

Berlin, Oktober 2016
Sarah Dahmann

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2SLS ..... Two-stage Least Squares

ATE ..... Average treatment effect

BAuA ..... *Bundesanstalt für Arbeitsschutz und Arbeitsmedizin* (Federal Institute for Occupational Safety and Health)

BKK ..... *Betriebskrankenkassen* (Company health insurance)

BMAS ..... *Bundesministerium für Arbeit und Soziales* (Federal Ministry of Labour and Social Affairs)

IV ..... Instrumental variable

KMK ..... *Kultusministerkonferenz* (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany)

LATE ..... Local average treatment effect

MCS ..... Mental Component Summary

MTE ..... Marginal Treatment Effects

NEPS ..... National Educational Panel Study

OECD ..... Organisation for Economic Co-operation and Development

OLS ..... Ordinary Least Squares

PCS ..... Physical Component Summary

PISA ..... Program for International Student Assessment

RKI ..... Robert Koch Institut

SF-12 ..... 12-Item Short-Form Health Survey

SHARE ..... Survey of Health, Ageing and Retirement in Europe

SOEP ..... Socio-Economic Panel

STEM ..... Science, Technology, Engineering and Mathematics

US ..... United States of America

WLE ..... Weighted maximum likelihood estimate

Chapter 1

Introduction

1.1 Motivation

"Intelligence plus character – that is the goal of true education."
(Martin Luther King, Jr.[1])

Education may come in very different shapes and facets. Formal education is delivered in particular institutions like schools and universities with a clear purpose, as for example graduation. In contrast, non-formal education takes place in an institutionalized context as well, but does not lead to a certificate. This includes extracurricular activities like learning an instrument and playing in the orchestra or being a member of a sports club. In addition, in many further contexts, informal education can take place, for example when parents interact with their children or when peers learn from each other. While this dissertation acknowledges that all these types of education are important in human development, it focuses exclusively on formal education.[2] Yet, to pay tribute to the broad definition, even solely within formal education, different aspects, including both qualitative and quantitative factors, are considered.[3]

Undoubtedly, (formal) education is one of the central determinants of not just individual success but also societal progress and economic growth. Many studies show that education is especially crucial for economic and financial prosperity of individuals, with the origins of this research dating back to Mincer (1974). Ever since, earnings are usually analyzed in Mincer-type regressions which explicitly incorporate the effect of education. One potential channel in this

[1] As stated in a Morehouse College Student Paper, The Maroon Tiger, in 1947 (King, 1947).
[2] In the following, the terms education and schooling are used interchangeably and refer to formal education.
[3] These are presented in more detail when discussing the measurement of education later in this section.

relationship is education generating human capital in individuals. This idea is embedded in the human capital model by Becker (1962)[4], who refers to activities that "influence future real income through the imbedding of resources in people" (p. 9), as investment in human capital. With schooling being an important such activity propagated by Becker (1962), this dissertation focuses on the effects of education on human capital. Still, Becker (1962) notes that not only schooling but also different kinds of investment may generate these benefits but they share the property of improving "the physical and mental abilities of people and thereby raise real income prospects" (p. 9).

Therefore, he marked the expression human capital as a broad term to define characteristics that enable an individual to perform labor, thus generating economic value. These characteristics, thus, include many different individual attributes which can range from pure knowledge over internalized traits like social skills, habits and personality, to personal conditions manifested in health. In its nature, human capital is not transferable from one individual to another, and is intangible. As a consequence, it is not straightforward to measure. Therefore, many empirical studies proxy human capital by characteristics which are easier to observe, like school enrollment, educational attainment or years of schooling (Wößmann, 2003). Especially the last, years of schooling, have been widely used as the single most important measure of human capital. However, when used *instead* of human capital, these education related proxies suffer from the severe disadvantage that the link *between* education and human capital cannot be uncovered. This, however, is of great interest as the theory of education *producing* human capital (human capital theory) stands in contrast to the hypothesis that education only *reflects* human capital due to self-selection into education based on ability (signaling theory, see Spence, 1973). Still, this distinction is crucial if one aims to identify true returns to education and to assess the potential for interventions to foster human capital production. This dissertation, therefore, explicitly distinguishes measures of human capital from education to analyze the causal link between the two.[5] More in particular, I focus on skills and health as two vital dimensions of human capital.

Skills are often distinguished further into cognitive (i.e. intelligence) and so-called non-cognitive skills.[6] The latter include personality traits, trust, preferences, values, and attitudes,

[4]While the static Becker model forms the basis of human capital theory, it has been overtaken by important extensions such as, for example, a dynamic dimension by Ben-Porath (1967).

[5]In this dissertation I exclusively focus on the causal direction of education impacting human capital. The reverse pattern, that higher skills or better health determine educational attainment is likely to co-exist but will not be considered. The estimation techniques applied account for this issue.

[6]The term skills is often used interchangeably with abilities in the existing literature. I do so too; however, I use skills in particular with reference to cognitive and non-cognitive skills.

amongst others, and are sometimes also referred to as soft or character skills.[7] Health, which forms another important part of an individual's human capital (Grossman, 1972a, 2000), also incorporates different dimensions, in particular physical and mental health.

Skills and health both are important determinants of economic outcomes, which forms the basis of their interpretation as human capital. For cognitive as well as non-cognitive skills, several studies establish for example positive effects on wages (Heckman et al., 2006a; Nyhus and Pons, 2005; Heineck and Anger, 2010). Thus, these skills are important determinants of individual labor market success. The same holds true for an individual's physical or mental constitution as impaired individuals are less able to work at their full potential with (long) periods of work absences as a consequence. Therefore, good health is a crucial determinant in an individual's ability to perform labor, and, hence, an important part of an individual's human capital (Grossman, 1972a, 2000). In addition to labor market success, these aspects of human capital have further benefits for a variety of other economic and social outcomes making them of crucial interest, not just for economists.[8] Thus, understanding the formation of these aspects of human capital is of particular importance.

In the remainder of this section, I first briefly elaborate on skill formation and health determination, and the potential role of education in these processes.[9] Next, I motivate why it is important to understand these human capital returns to education, both from an individual's and a policy maker's perspective. Finally, I extensively discuss the measurement of education in these analyses and argue why I include further aspects beyond the classical indicator of years of schooling. In the following sections, the particular contributions of this dissertation are highlighted (Section 1.2) and a detailed overview of the following chapters is given (Section 1.3).

Human Capital Formation. In the economic literature, the formation of cognitive and non-cognitive skills is often modeled as in Cunha and Heckman (2007), who assume that skills persist over time, that skills beget skills through a multiplier process, and that they change due to environmental influences. Next to genetics, referred to as the channel of *nature*, which determine initial endowments, investments by parents and schools can influence skills, referred to as

[7] The expression non-cognitive is not very accurate as these skills clearly also entail cognitive components. Yet, for consistency within the existing literature I use the term non-cognitive skills as a mere distinction to cognitive skills in the strict sense. See Almlund et al. (2011) for details.

[8] Examples include, but are not limited to, further individual outcomes including education and performance (Coneus et al., 2011; Piatek and Pinger, 2016; Duckworth and Seligman, 2005), intergenerational spill-over effects (Bhalotra and Rawlings, 2011; Coneus and Spiess, 2012), and macro-level outcomes (Hanushek and Wößmann, 2008). Moreover, research shows that e.g. personality traits are at least as important as cognitive skills in determining social outcomes, including criminal behavior, marital stability, as well as health and mortality (Heckman et al., 2013; Roberts et al., 2007).

[9] The underlying theoretical models are presented in the following chapters in more detail.

nurture.[10] While most likely both mechanisms interact, the former leaves policy makers largely unable to act whereas the latter provides the opportunity to target skills, and is, therefore, of most interest to researchers. Furthermore, the productivity of investment in skills is assumed to increase with the existing stock of skills. This multiplier effect, therefore, suggests that resources allocated to targeting individuals during early stages of life are most efficient. Indeed, studies show that early interventions, like the Perry Preschool Project, improve both cognitive and non-cognitive skills, where especially changes in personality traits account for later-life benefits as cognition changes only in the short-run (Heckman et al., 2013).[11] Similarly, although health is assumed to be pre-determined in large parts, due to genetic and prenatal influences, Grossman (1972a, 2000) explicitly models it as a function of investment in addition to the existing stock of health. Again, early life experiences may have long-term benefits, as several studies establish a protective effect of schooling; in particular on physical health in adulthood (Kemptner et al., 2011; Lleras-Muney, 2005; Kenkel et al., 2006). Hence, for both dimensions of human capital, skills and health, the pre-educational and educational period in life seem to be critical. The schools and institutions that children and adolescents attend may, therefore, be important in determining their later life human capital.

Human Capital Returns to Education. Understanding these intangible effects of education in more detail is important from the perspective of both the individual, or his and her parents respectively, and the policy maker. Individuals have to make decisions on opting out of or continuing education at several times in their life course. While continuing education is often costly in the short-term due to forgone earnings, tuition fees, or effort, it only pays off monetarily after completion in terms of e.g. higher wages or reduced unemployment risk. If, and only if, these benefits outweigh the incurred costs, it is rational for the individual to opt for staying in education. Of course these costs and benefits may differ between individuals as they may value certain aspects more or less, i.e. differ in their utility functions. Some of these aspects can be monetized by e.g. the amount they would pay off in terms of wages, like a higher intelligence. Others that cannot be monetized directly but are important for the individual's quality of life, like health, often can be expressed by the amount of money people would be willing to give up in exchange. In both cases, however, these intangible returns to education are important potential benefits that should be taken into account by the individual decision maker. More accurate knowledge on these returns is, therefore, indispensable for an individual to arrive at an optimal decision.

[10] Initial endowments are often used interchangeably with the stock of human capital at birth. But strictly speaking, the human capital at birth is already subject to nurture through in-utero influences during pregnancy.

[11] Further examples and a more extensive overview of the existing literature can be found in the respective chapters of this dissertation.

Additionally, this knowledge is also crucial from a policy perspective. Understanding how and to what extent human capital is formed by education, helps policy makers to target, for example, skills or to improve health in the population through educational interventions. This can serve multiple purposes, including efficiency gains and fairness arguments. Efficiency refers to the optimal allocation of the available resources to achieve a specific outcome, like improving a society's welfare. Since the resources are limited, education usually cannot be tailor-made for every individual. This involves the risk that some individuals who are not adequately supported may not develop to their full potential. This is problematic if a society's welfare is based on innovations and progress. In this case, the waste of potential may be detrimental. In turn, a minimum level of education may be as important for overall welfare. Hence, in any case, policy makers need to adequately prioritize their aims and weigh the advantages and disadvantages of different schemes. When designing educational policies, they should, much like individuals, also weigh costs and benefits of their actions and decisions to arrive at an efficient distribution of the limited resources. In this context, intangible effects of education may constitute important costs and benefits. Although they are hidden at first glance, they have important consequences for e.g. the labor market or the health care system, and therefore must be considered. In addition, this decision making is a constantly evolving process. New insights into educational processes, a change in institutional frameworks and conditions or current political interests may challenge the status quo. In this case, a reevaluation may suggest an educational change to be necessary to continue an efficient use of resources, especially in the long-run. Although frequently costly, as more input is required, educational reforms may pay off in the long-run if for example better educated students lead to higher productivity on the labor market or reduced health care costs. Again, to assess whether the benefits of such a reform indeed outweigh the costs, policy makers need a thorough evaluation and in principle monetize all induced effects, including potential returns on human capital.

From a fairness perspective, in turn, more detailed analyses of subgroups shed light on how to target disadvantaged individuals. This helps overcoming or avoiding rising inequality as human capital is an important channel in the intergenerational transmission of socio-economic status. Educational attainment and labor market success of children is highly correlated with that of their parents (see e.g. Black and Devereux, 2011, for an overview on intergenerational mobility). In the literature at least two channels can be identified of how this correlation can be transmitted through human capital. On the one hand, skills may have differential effects on educational attainment and labor market success, depending on the socio-economic background. For example, using data from the United States of America (US) on 24 to 32 year-olds,

Lundberg (2013) shows that the returns to non-cognitive skills vary substantially depending upon maternal education: while for more privileged students a high degree of conscientiousness positively correlates with the probability of college graduation, this trait yields no significant returns for men with lower maternal education. In contrast, openness to new experiences is an important correlate with educational success among disadvantaged students only.[12] The influence of cognitive skills, in turn, shows little heterogeneity between the two groups. Therefore, particularly non-cognitive skills show potential to specifically target children from disadvantaged backgrounds. On the other hand, human capital can be directly passed on from parents to their children, thereby influencing later life outcomes. This transmission can occur either directly through genetic inheritance (nature) or through upbringing (nurture), or through an interaction of both. For policy makers particularly the channel of nurture offers scope for action, which reinforces the importance of studying the potential of education in skill formation. Empirically, it is not easy to disentangle the influences of nature and nurture, but is possible with data on twins or with the distinction of parents into biological and social parents in case of adopted children. In any case, the literature agrees that it is the interaction of both that leads to the transmission of human capabilities from parents to children. For example, Anger and Heineck (2010) show that there is a high correlation between cognitive skills of parents and their children in Germany, but that this transmission likely occurs not purely genetically but also through parental investment as skills related to past learning are transmitted stronger than those based on innate abilities. Anger (2012) shows that the strong intergenerational correlation extends from cognitive to non-cognitive skills. Yet, there is a large discrepancy in the extent between different types of skills: Cognitive skills correlate much more strongly between the two generations than it does for non-cognitive skills. The latter, therefore, provide much larger scope for interventions. The importance of both, nature and nurture, is also evident in studies based on adoptive children attempting to quantify both (e.g. Björklund et al., 2007). According to Plug and Vijverberg (2003) close to half of an individual's ability is influenced through nurture, which should be exploited by policy measures to guarantee equality of opportunity.

Further, it is not only skills that are transmitted from one generation to the other, but also health is influenced directly by parental characteristics. For Germany, Coneus and Spiess (2012) find a positive relationship between maternal and child health in the first three years of life. These findings are confirmed by Bhalotra and Rawlings (2013) for developing countries, which in addition reveal that it is not only the mother's contemporary health that matters for intergenerational transmission, but also her own childhood health. This suggests that ef-

[12] In Germany, this pattern is similar, where only among students from non-academic backgrounds study intentions increase with their openness to new experiences (Peter and Storck, 2015).

fects persist. In addition, parental education is shown to improve child health outcomes, for example in the US (Currie and Moretti, 2003), but also in countries with almost universal health insurance coverage like Germany (Kemptner and Marcus, 2013). Case et al. (2005) and Currie (2009) suggest focusing more on health as a driver of intergenerational transmission of socio-economic status due to the importance of childhood health for future educational and labor market outcomes. Also for Germany, Salm and Schunk (2012) find that disparities in the prevalence of health conditions, particularly mental disorders, account for a substantial part of the gap in child development between socio-economic groups. Therefore, health is a further important channel in the intergenerational persistence of socio-economic status, in addition to skills.

Hence, from the perspective of individuals and policy makers, while also considering both efficiency and fairness arguments, it is crucial to identify the potential of education in the formation of human capital. For these reasons, this dissertation provides evidence on different intangible returns to education. However, it differs from the classical economic approach as it goes beyond a unidimensional definition of education.

Measuring Education. In the empirical economics literature, by far the single most common measure of education is years of schooling (Hanushek and Wößmann, 2008). This measure dates back to Mincer (1974) and has the advantage of being easily observed. Furthermore, it is a clearly ordered and linear metric that can be used across countries, thus easing international comparison. Yet, it can be heavily criticized for several reasons.

First, a technical issue questioning the increasing monotonic nature of the measure is that *actual* years of schooling may depend on the learning capacity of the individual and, therefore, not adequately mimic educational attainment. This is the case, for example, with grade repeaters for whom this additional year of schooling includes only repetition, and hence, is not the same as an extra year of education for non-repeaters. In addition, especially in tertiary education where completion time is individual specific, more able individuals who learn faster may run more quickly and efficiently through the educational system (Schneider, 2010). A simple solution to this problem is to use *hypothetical* years of schooling instead, which count the years that are usually necessary to complete the degree the person obtained. It comes at a disadvantage though because it does not consider years of education that do not result in a degree because the person drops out.

Second, using years of schooling in economic analyses rests on the assumption of linearity. This measure implicitly assumes that an additional year has the same effect regardless of whether it is the individual's first year of education or his or her last (Wößmann, 2003). In par-

ticular, with respect to human capital, this may be inadequate. The first year may yield much higher returns if, for example, basic skills passing a certain threshold add most value. This is, for example, reflected in earnings, which exhibit diminishing returns with increasing levels of education (Psacharopoulos, 1994). Wößmann (2003), therefore, argues that, depending on the years of schooling already accumulated by a person, an additional year of schooling should be weighted differently.

Third, and very similarly, the proportional interpretation of coefficients can be criticized. A person with two years of schooling may not necessarily have received twice the education that a person with only one year of schooling received, if content is disregarded.

This leads to the fourth and probably most important critique: using years of schooling as a measure for education completely disregards any quality aspects of education. One year of schooling is assumed to have the same impact regardless of the field of study, of the curriculum, of the quality of teaching, of the educational infrastructure or of the efficiency of the education system (Mulligan and Sala-i Martin, 2000; Wößmann, 2003). But not all time spent at school may be effective in transmitting education. For example, Hanushek and Wößmann (2008) note that schooling that does not improve cognitive skills is of limited importance for economic outcomes.[13] Therefore, especially when comparing education across countries, the same number of years of schooling may still not imply the same level of education. Wößmann (2003), therefore, argues that years of schooling should be weighted differently in accordance with the country's quality of the education system. However, this aspect is not only problematic in cross-country comparisons, but also in within-country analyses. In particular, in highly stratified educational systems, years of schooling may not adequately reflect the educational content. In the US, within-school tracking is prevalent, such that peers at school-wide level and the demographic characteristics are similar for both lower and higher achieving students and mobility between tracks is more common than elsewhere (Betts, 2011). In such systems years of schooling may be more valid as a measure for education. In contrast, in many European countries, including Germany, the educational system relies on different tracks which are usually offered at separate schools. Decisions on the educational path, therefore, often have to be taken rather early and are harder to reverse later. The resulting degrees often come with a fixed number of years, such that the choice of education is not subject to yearly change but often remains fixed for several years. Most importantly, though, the tracks differ substantially in the educational level provided and the ability required. The German educational system comprises three basic types

[13] Therefore, other studies (e.g. Jamison et al., 2007) use international student achievement tests to measure the quality of education. But since these are rather a proxy of acquired human capital, this does not permit studying the relationship *between* education and human capital.

of secondary school: after usually four years at elementary school, students choose between the basic track, *Hauptschule*, the intermediate track, *Realschule*, and the upper track, *Gymnasium* (henceforth referred to as academic-track or simply high school). After graduation from the basic or the intermediate track, students usually start an apprenticeship or vocational training. Graduation from the highest track leads to the *Abitur*, the university entrance qualification. Hence, the educational level between these three tracks and the content delivered clearly differ. Therefore, it is inadequate to simply count years at school, regardless of whether they are spent at basic track school or at high school.

Hence, despite its convenience, solely using the number of years at school as a measure of education can be misleading in policy debates (Hanushek and Wößmann, 2008).[14] This dissertation acknowledges this multidimensionality of education, considering both qualitative and quantitative aspects as educational inputs in the formation of human capital. Each of the following chapters sheds light on different such factors, complementing each other with respect to both measuring education as well as human capital. Chapter 2 and 3 investigate curricular and efficiency aspects. More specifically, they consider an increase in weekly class hours, which results from compressing a larger curriculum into the same number of years. This leads both to a higher learning intensity and a larger amount of the curriculum covered at a particular point in time. Since this change was only implemented in high schools, both chapters focus exclusively on students attending this track. This also avoids confusing the level of education as only students of one single track are considered. In addition, Chapter 3 investigates whether the efficiency gains, when compressing the same curriculum into fewer years, are translated into a higher efficiency in learning such that the same competences can be acquired in shorter time. Specifically in this case, the necessity for an alternative measure to years of schooling becomes evident as the years can vary despite the educational content being completely maintained. Furthermore, this chapter, which focuses on cognitive skill acquisition, nicely illustrates the importance of considering further aspects: Most related studies in the existing literature identify beneficial effects of an additional year of schooling without any hint at whether these also rely on content or contextual changes. In contrast, this chapter of this dissertation provides evidence on the underlying mechanisms, thus providing clearer indications for policy makers. In contrast to the aforementioned two chapters, Chapter 4 turns to the classical quantitative measure of education and employs years of schooling. This is advantageous because it facilitates comparing

[14]Despite these shortcomings, researchers still argue, that years of schooling are an important indicator of education, in particular with its advantages for international comparisons (Pfeiffer et al., 2015). It should, therefore, continue to be developed further and improved. Educational degrees as a categorical variable may sometimes be used as an alternative. This measure, however, can bear difficulties in international comparisons and shares some of the disadvantages with the years of schooling measure such as, for example, disregarding any education ultimately not resulting in a degree.

its results to those of other studies. In particular, in the context of mental health, which is the focus of this chapter, it is of interest to contrast results to effects on physical health in existing studies. To deal with the aforementioned critique on this measure, hypothetical years are used instead of actual years of schooling. Furthermore, potential non-linearity in effects is specifically accounted for by estimating local effects at different parts of the educational distribution. As such, the estimated effect of an additional year of schooling may differ for persons accumulating only few years of schooling compared to those with a large number of years of schooling. In addition, effect heterogeneity is explicitly modeled along the intrinsic preference of individuals for obtaining education. This allows investigating how the impact of obtaining at least median years of education differs between individuals with a relatively low and a relatively high distaste for education.

1.2 Contributions

Each of the three following chapters of this dissertation delivers an independent contribution to the empirical literature by investigating a specific research question. However, they also share five common contributions.

First, each chapter identifies intangible effects of education. These are fundamentally different from tangible measures, which are uniformly quantifiable, like wages or unemployment risk. While these tangible effects have historically received much more attention, especially among economists, the integration of psychological measures into large survey data sets since the new millennium triggered the shift of this focus. However, interpretation of effects may be even harder in the case of an ordinal instead of a cardinal ranking.[15] The existing literature on the causal determinants of these intangible outcomes is naturally only recent and, hence, still limited with many open research questions. All chapters, therefore, deliver important and new insights into these processes. More in particular, this dissertation provides first evidence on the malleability of personality traits in adolescence through a non-targeted large-scale intervention. It also successfully disentangles potentially important mechanisms through which education may influence cognitive skills and investigates their role in gender inequalities. And it significantly complements the literature on mental health, so far limited to disorder-specific diagnoses at the extensive margin, by a generic measure of mental health which is able to capture changes at the intensive margin and is applicable to the general population.

[15] For example, this is the case with personality traits, where high values on a trait may not necessarily go along with social desirability. Often, less extreme values can be optimal, or certain traits like a high agreeableness may be beneficial for e.g. social relations but detrimental in wage negotiations. In contrast, cognitive skills and mental health, as used in this dissertation, are measured on cardinal scales.

Second, all chapters investigate education in a country where it is cheap and easy to access. Focusing on Germany, the chapters complement the largely US-focused literature with evidence on a modern country that has a different institutional framework. On the one hand, the overwhelming majority of schools and universities in Germany are publicly funded and free of charge, thus easing access. On the other hand, however, an early tracking system is in place, usually separating students after grade four, thus making education in Germany highly stratified. Despite equal access, differences in educational attainment at the early stage may, therefore, be even harder to overcome later. Hence, the role of education in processes of human capital formation and the development of inequality can be different from that in other countries. Additionally, educational policies are at the discretion of each one of the sixteen *Bundesländer*, the federal states, which yields itself again differences in education regulations across the country. This regional variation is crucial for the credible identification of causal effects. Policy changes cannot only be exploited as exogenous variations in education in simple before and after comparisons. In addition, they allow for state-of the-art microeconometric techniques if these changes are administered at different times in the different states. In this case, time effects can be specifically modeled such that any other potential influential factors can be accounted for. This is particularly relevant if further policy changes occur at the national level which may affect the outcomes of interest, like, for example, a change in health policy.

Third, and directly related to the above aspect, all chapters aim at investigating causal effects. In the empirical literature, the fundamental problem in identifying causality lies in that the same person cannot be observed at the same time, both with and without treatment. The term treatment originates from experimental studies, often applied in medical research, and refers to the variable, whose effect we are interested in, i.e. education in this case. Much like in the experimental literature, a randomized allocation to a treatment group, i.e. persons with more education, and a control group, i.e. persons with less education, would be ideal. This way, self-selection into education could be circumvented as both groups should be comparable in terms of their observed and unobserved characteristics due to the randomization. However, in the real world, individuals make their own educational choices and these choices may inherently depend on the outcome of interest or unobservable factors that are correlated with both education and the outcome. This, what is termed endogeneity, leads to the treatment and the control group no longer being comparable and for simple correlation estimates to be biased. As experiments, however, are unfeasible or unethical in many real-world contexts, different microeconometric techniques can be applied to mimic the idea behind experiments with real-world data. Among others, this can be done by exploiting exogenous shocks, which are

beyond the individual's control, on the education obtained. If these shocks happen to some individuals, but not others, the allocation to treatment is close to random. Crucial is, however, that the determinants of whether an individual is affected by the shock are beyond their own control; for example, their date of birth. Well suited candidates for such exogenous shocks are, therefore, any reforms applying only to persons born after a certain cut-off date, but also regional variation in case choice of residence is unrelated to the treatment, or policy trends in e.g. the expansion of education.

Fourth, each of the chapters relies on changes in educational policies for the identification of causal effects. Chapter 2 and 3 investigate a reduction in the length of high school by one year that was implemented between 2001 and 2007. Chapter 4 exploits the extension of compulsory schooling in former West Germany implemented between 1949 and 1969, as well the increasing supply of higher education through the opening of new universities in the twentieth century. Implicitly, each analysis also constitutes a policy evaluation of the respective change with respect to intangible effects on human capital.

Fifth, each chapter goes beyond the specific research question to shed light on important heterogeneities and potential mechanisms that explain the results. Chapter 2 investigates whether effects differ by individual characteristics to infer on consequences for socio-economic inequality, while Chapter 3 focuses in particular on gender differences. In Chapter 4 differential effects along the educational distribution are the focal point: effects are estimated separately for individuals obtaining presumably different levels of education, and, in addition, effect heterogeneity with respect to the individual preference for education is explicitly modeled. Furthermore, all chapters exploit the rich information contained in the survey data sets that are used in this dissertation to augment the analyses by important channels: (i) Behavior, which includes engaging in leisure-time activities as music, sports, and reading and, thus, resembles non-formal and informal education. (ii) Satisfaction, including emotions, which captures the self-perception of estimated changes and entails valuable information on the psychological consequences. (iii) Support, which reflects parental involvement with their children's education and additional resources such as paid tutor lessons. (iv) Medical diagnoses of disorders, which provide a more specific insight into an individual's mental health condition. Together, these aspects constitute important channels and provide a more detailed analysis of how education ultimately impacts human capital.

In addition to these common contributions, the chapters complement each other well in three dimensions. First, as already elaborated in Section 1.1, each chapter looks at different aspects of education, together considering both qualitative measures – including the curricu-

lum, learning intensity, and timing (Chapters 2 and 3) – as well as quantitative factors – like years of schooling (Chapter 4). Second, while all chapters estimate intangible effects of education, each investigates a specific dimension of human capital, overall addressing non-cognitive skills, cognitive skills, and health: Chapter 2 looks at personality traits, Chapter 3 investigates crystallized and fluid intelligence, and Chapter 4 analyzes mental health. Third, the chapters complement each other from a life-course perspective as they focus on samples of different age groups: together they address seventeen-year old adolescents still enrolled in secondary education (Chapters 2 and 3), young adults at the time of graduating from high school (Chapter 3), and adults between the ages 50 and 85, long after the completion of secondary education (Chapter 4).

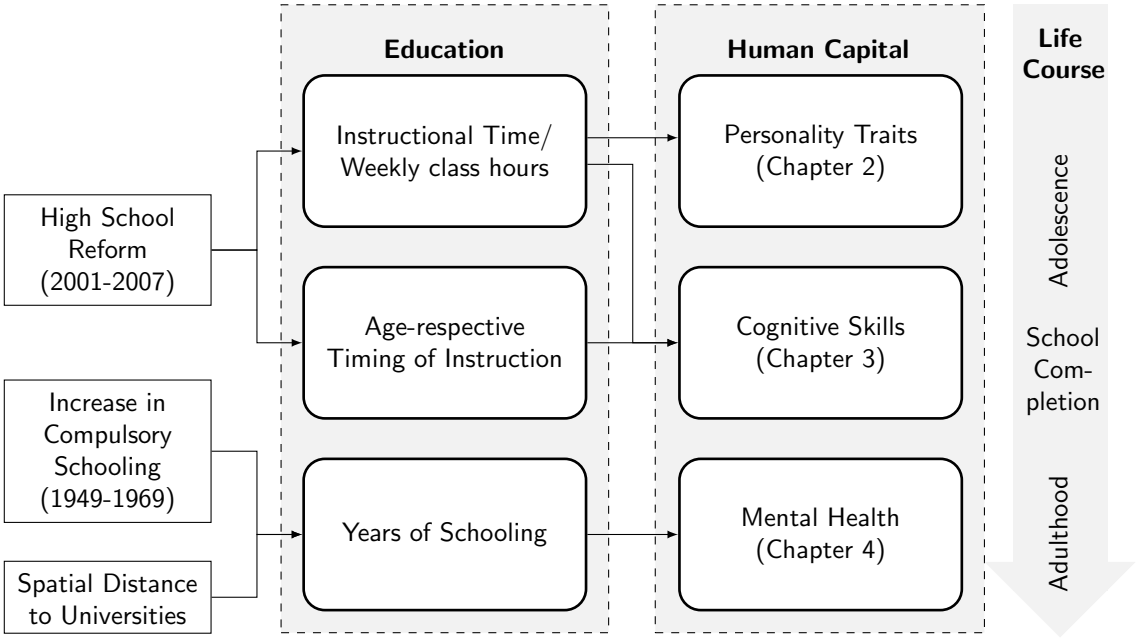
1.3 Overview and Summary

This dissertation is based on three independent research articles that are ordered along the life-course and constitute Chapters 2, 3, and 4 of this dissertation. They all connect in estimating returns to education on human capital outcomes. Figure 1.1 illustrates the specific aspects of education analyzed together with the different dimensions of human capital. Additionally, the figure depicts the respective exogenous sources of variation exploited in the analyses for causal identification (at the left) and the different life phases of the samples considered (at the right).

Table 1.1 provides a more detailed overview of all chapters including their research question, the data and methodology used, as well as collaborations with co-authors and publication output. In the following, each chapter is described in more detail summarizing the main findings.

Chapter 2 investigates the short-term effects of a reduction in the length of high school on students' personality traits using a school reform carried out at the state level in Germany as a quasi-natural experiment. Starting in 2001, academic-track high school education was reduced by one year in most of Germany's federal states, with the overall curriculum left unchanged. This enabled students to obtain their university entrance qualification after only 12 rather than 13 years of schooling, but it increased learning intensity through an increase in weekly class hours. Exploiting the variation in the length of high school over time and across states, the effect of learning intensity on students' Big Five personality traits and on their locus of control is identified. Using representative data on adolescents and young adults from the German Socio-Economic Panel (SOEP) study, the estimates show that shortening high school caused students to be less emotionally stable. Furthermore, the results point to important heterogeneous effects. In addition to regional and gender differences, students not living with

Figure 1.1: Structure of Dissertation



Source: Own illustration.

both parents and students with migration background showed stronger personality changes following the reform: they became more open and extroverted, and more open and conscientious, respectively. Thus, the chapter concludes that the educational system plays an important role in shaping adolescents' personality traits. This chapter is the first study to investigate a nationwide large-scale reform to analyze the causal effect of education on personality which can be credibly estimated, it is the first to give important insights into heterogeneous effects and potential mechanisms, and it complements the largely US based literature and the literature focusing primarily on young children.

Chapter 3 focuses on cognitive skills and investigates two mechanisms through which they may be affected by education. I rely on the same reform analyzed in Chapter 2 for causal identification but conduct two separate analyses: 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 SOEP data on seventeen year-olds. Second, I investigate the influence of earlier instruction at younger ages, using data from the German National Educational Panel Study (NEPS) 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 re-

Table 1.1: Overview of Chapters

	Chapter 2	Chapter 3	Chapter 4
Title	The Impact of Education on Personality – Evidence from a German High School Reform	How Does Education Improve Cognitive Skills? Instructional Time versus Timing of Instruction	Is There a Protective Effect of Education on Mental Health?
Main Findings	Learning intensity decreases students' emotional stability; strong heterogeneous effects by student characteristics on other personality traits	Instructional time increases male students' cognitive skills and widens gender skill gap	Education has no causal protective effect on mental health
Data	SOEP	SOEP, NEPS	SOEP
Identification Method	Difference-in-differences Estimation	Before/After comparison of a quasi-natural experiment, Difference-in-differences Estimation	Instrumental Variables Estimation, Marginal Treatment Effects
Co-author	Silke Anger	–	Daniel Schnitzlein
Discussion Papers	IZA Discussion Paper No. 8139, SOEPpapers 658, IAB-Discussion Paper 29/2014, CINCH working paper 2015/05	SOEPpapers 769, CINCH working paper 2015/04	–
Awards	Shortlisted for BeNA Innovative Research Award 2014 (Berlin Network of Labor Market Research)	Sir Alec Cairncross Prize 2016 (best paper by a young economist, Scottish Economic Society)	–

veal that increasing instructional time aggravates gender differences in numeracy. This chapter is the first study to disentangle both mechanisms, giving important policy insights, it complements the existing literature on Germany, which was contradicting international evidence, it includes females to specifically understand gender differences, and it differentiates dimensions of cognitive skills.

Chapter 4 analyzes whether education has a protective effect on mental health. To estimate causal effects, an instrumental variable (IV) technique is implemented with two different instruments to estimate local average treatment effects (LATE) at different parts of the educational distribution: (i) a reform extending compulsory schooling by one year implemented in West German federal states between 1949 and 1969 and (ii) the individual availability of higher education measured by the spatial distance to the nearest university at age 19. The analyses are

based on rich individual SOEP data on adults aged 50 to 85, augmented by detailed information on universities from the German Rectors' Conference. Analyses on the Mental Component Summary (MCS) score as a generic measure of overall mental health are complemented by disorder-specific diagnoses. Results support existing evidence on a positive relationship between completed years of secondary schooling and mental health in standard Ordinary Least Squares (OLS) estimations. However, the IV estimations reveal no such causal protective effect. If any, the estimates point towards a negative effect among the lower educated. These results are confirmed when explicitly modeling effect heterogeneity through marginal treatment effects (MTE). This chapter provides first evidence on the causal effect of secondary education on mental health in Germany, it employs a generic measure of overall mental health to complement existing studies limited to specific mental disorders, it considers potential heterogeneity in effects, and it adds to the empirical literature on estimating MTE which has emerged only recently.

Finally, Chapter 5 concludes with a critical discussion of the aforementioned chapters, policy implications, and suggestions for further research.

Chapter 2

The Impact of Education on Personality

Evidence from a German High School Reform*

2.1 Introduction

In recent years, a growing body of literature has emerged on the importance of personality traits as determinants of individual economic and social outcomes.[16] Studies investigating the impact of personality on labor market outcomes show that certain characteristics such as emotional stability are rewarded with higher wages, while characteristics such as agreeableness are penalized in the labor market (Heckman et al., 2006a; Nyhus and Pons, 2005; Heineck and Anger, 2010). Furthermore, personality has been shown to affect labor market success through occupational sorting (John and Thomsen, 2014) and job search behavior (Caliendo et al., 2015). Not only has personality been linked to specific labor market outcomes; it has also been found to affect educational success, as shown by studies on school dropout probability (Coneus et al., 2011), educational attainment (Piatek and Pinger, 2016), and academic performance (Duckworth and Seligman, 2005). The returns to personality may also differ, as Lundberg (2013) shows that

*This chapter is based on joint work with Silke Anger. We thank Alexandra Avdeenko, Deborah Cobb-Clark, Nabanita Datta Gupta, Thomas Dohmen, Adrian Hille, Daniel Kamhöfer, Peter Kuhn, Philip Oreopoulos, Erik Plug, Mari Rege, David Richter, Thomas Siedler, and seminar and conference participants at DIW Berlin, Maastricht University, NIW Hannover, University of Konstanz, University of Nuremberg, the Berlin Network of Labor Market Research, the AMCIS Conference "Education Systems", the Spring Meeting of Young Economists, the Annual Conference of the Scottish Economic Society, the RWI Research Network Conference, the Annual Conference of the European Society for Population Economics, the Annual Conference of the German Economic Association, the International Workshop of Applied Economics of Education, the Annual Conference of the European Association of Labour Economists, the Essen Health Conference, and the Summer School on Socio-Economic Inequality for helpful comments and discussions.

[16] For a detailed overview, see Almlund et al. (2011).

college graduation requires students from lower socio-economic backgrounds to be open towards new experiences while others need to be more conscious. Moreover, research shows that personality traits are at least as important as cognitive skills in determining social outcomes such as criminal behavior, marital stability, and health and mortality (Heckman et al., 2013; Roberts et al., 2007).

While the association between personality traits and economic and social outcomes is widely acknowledged, there is far less research in economics on the factors that affect the formation of these non-cognitive skills. The literature has identified two main channels that shape non-cognitive skills: *nature*, referring to the fact that personality is strongly heritable, and *nurture*, encompassing all environmental factors such as culture, social factors, family background, and individual or situational factors which may influence personality. In their model of skill formation, Cunha and Heckman (2007) argue that it is the interaction of these mechanisms – nature *and* nurture – that determines skill formation, but that the two channels cannot be disentangled. Their model suggests that the development of skills takes place especially during the early (pre-)educational period in life. This is consistent with the psychology literature, according to which personality traits develop mainly during childhood and adolescence, and remain relatively stable later in life (Costa and McCrae, 1994). As a nurturing factor, education during childhood and adolescence may, therefore, constitute a critical determinant in an individual's long-term formation of personality. Yet, there is little direct evidence on the effect of schooling on personality traits. Most of the research focuses on interventions targeting children at a relatively young age, such as the Perry Preschool Program, which is aimed at three- to four-year-old preschoolers (Heckman et al., 2013), and the Project STAR (Dee and West, 2008).

In the present study, we focus on the impact of education on personality traits in adolescence. Our research question is threefold: First, we assess whether schooling, and more specifically learning intensity affects adolescents' personalities. Second, we examine heterogeneous effects by analyzing which students are most likely to exhibit changes in personality following changes in the educational system. And finally, we investigate the underlying mechanisms by disentangling various potential channels of impact.

To address these questions, we use a nationwide educational reform carried out in Germany in the 2000s as an exogenous variation in schooling. The reform shortened academic-track high school by one year, leaving the overall curriculum unchanged, and hence reduced the total number of years to obtain a high school degree from 13 to 12. We exploit the variation over time and across federal states to analyze causal effects of learning intensity on students'

personality traits, which are measured by the dimensions of the Big Five personality inventory (Openness, Conscientiousness, Agreeableness, Extraversion, Neuroticism), and locus of control. Using data on adolescents and young adults from the German Socio-Economic Panel (SOEP) study, we find that the reduction in the length of high school by one year led to considerably lower emotional stability among high school students. Moreover, our estimates point to important heterogeneous effects. While this decrease in emotional stability was more pronounced among students in East Germany, West German students and students with better educated parents faced an increase in agreeableness. Female students in particular seemed to feel less in control over their lives. In contrast, disadvantaged students from either non-intact or migrant families benefited substantially in terms of openness and extraversion, and openness and conscientiousness, respectively.

We find suggestive evidence that the reform effect is driven by adverse effects on health, school performance and emotions, and by stronger student-teacher or student-student interactions as a result of longer school days, and not by a change in leisure-time activities. We conclude that the educational system plays an important role in shaping adolescents' personality traits.

Our study adds to the existing literature in several ways. First, to the best of our knowledge, this is the first representative study that exploits a nationwide change in the educational system as quasi-natural experiment to analyze the causal effect of education on personality. In contrast to previous studies on the impact of schooling on the malleability of personality traits in young children, our treatment is large-scale and not locally restricted, as it affects students from almost an entire country. Second, to identify the causal effects of education, we exploit variation in schooling over time and across states. The high school reform has been gradually introduced in almost all of Germany's federal states. Hence, we are able to isolate the causal effect of learning intensity from any other potential influential factors or policy changes. Furthermore, we reduce the risk that potential unobserved effects bias our estimates by including students who graduated several years before and after the reform, which enables us to establish a long-lasting effect of the reform rather than an artifact of its implementation. Third, our data provide rich information on individuals' socio-economic backgrounds, which allows us to examine whether the personalities of particular groups of students were more malleable. Pre-reform characteristics of students and their families such as migration background, parental education, or resources may affect the ability to cope with reform-induced changes. As a consequence, different students may react differently to the high school reform in terms of both the type of personality traits affected and the extent to which personality traits change. Fourth, by

examining the malleability of personality in adolescence, we complement previous studies on skill formation that focused primarily on young children. Evidence on whether personality is malleable even in adolescence is crucial for policy makers who may want to target individuals who are too old for childhood intervention programs. In addition, educational policy makers should be informed about potential second-order effects on students when implementing school reforms. Finally, we complement the previous largely US-based research with evidence on a modern country with a different institutional framework. Differences in educational settings between the US and other countries could lead to different processes of skill formation, in which education, family factors, and other determinants of personality may play different roles. Hence, cross-country comparisons may be meaningful to understand the importance of educational settings for non-cognitive skill formation.

The next section presents the theoretical background on skill formation and gives an overview of previous research. Section 2.3 explains the reform and discusses potential effects on adolescents' personality traits. Section 2.4 describes the data and the empirical strategy, and Section 2.5 presents the results. Section 2.6 elaborates on several robustness checks. Section 2.7 concludes and discusses the implications of our findings.

2.2 Theoretical Background and Previous Literature

2.2.1 Skill Formation Model

Personality traits are a significant part of an individual's non-cognitive skills.[17] Both cognitive abilities and non-cognitive skills constitute personal skills, which belong to an individual's overall human capital. A prominent approach to describe the formation of such skills has been developed by Cunha and Heckman (2007). They propose that an individual's present skill stock depends on his or her past skill stock, previous investments, and environmental characteristics, according to the following model:

$$\theta_{t+1} = f_t(\theta_t, I_t, h) \tag{2.1}$$

where a vector of skill stocks at age $t+1$, θ_{t+1}, depends in some functional form $f(\cdot)$ on the past vector of skill stocks (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*. The former, self-productivity, occurs when skills persist such that higher

[17] Examples of further non-cognitive skills are trust and (e.g., time or risk) preferences (see, e.g., Almlund et al., 2011).

skills in one period create higher skills in the subsequent period. This is not restricted to one and the same skill, but also applies to cross effects between different skills, and is the case whenever $\partial f_t(\theta_t, I_t, h)/\partial\theta_t > 0$. The latter, dynamic complementarity, manifests that the productivity of an investment is increasing with higher existing skills and occurs whenever $\partial^2 f_t(\theta_t, I_t, h)/\partial\theta_t\partial I_t' > 0$. Cunha and Heckman (2008) test and verify both propositions empirically. The resulting multiplier effect suggests that investments are most productive in early stages in life, which implies that childhood constitutes a bottleneck period for skill formation. Hence, education during childhood may constitute a critical determinant in an individual's long-term formation of personality.

2.2.2 Previous Empirical Studies

In line with this theory, Fletcher and Schurer (2015) show that (adverse) childhood experiences have high predictive power in explaining personality traits at age thirty. Several intervention studies have therefore been targeted at children of preschool age. Even though most of these do not focus primarily on improving personality traits, they make it possible to study the impact of education on personality. A prominent example is the Perry Preschool Program, which provided extra classes and teacher visits at home to three- to four-year-olds. The program led to large benefits in later-life outcomes, such as improved labor market outcomes and a reduction in criminal incidence (Schweinhart et al., 2005). Though the program did not focus primarily on improving personality traits, Heckman et al. (2013) show that the reported benefits are fully attributable to a permanent improvement in non-cognitive skills, while there has not been any lasting effect on cognitive abilities. The project STAR revealed similar results; children who were randomly assigned to small kindergarten classrooms showed positive changes in personality, assessed by teacher-reviews measured in fourth grade (Dee and West, 2008). Other interventions, such as the PATHS program, focused more explicitly on improving non-cognitive skills and showed that personality traits are malleable during childhood (Bierman et al., 2010).

While Cunha and Heckman (2008) find a strong multiplier effect of early investments, their empirical results also reveal that the critical period of investment lasts longer in life for non-cognitive skills than for cognitive skills. Yet, the evidence on malleability of personality traits in adolescence remains limited. Martins (2010) shows that a program targeted at 13- to 15-year-olds in Portugal improved student achievement by increasing motivation, self-esteem, and study skills. Furthermore, the National Guard Youth ChalleNGe program in the United States increased discipline and emotional stability among high school dropouts (Bloom et al., 2009). Ex-

ploiting changes in the compulsory minimum school leaving age in Australia, Li and Powdthavee (2014) find that an increased number of years of schooling raises individuals' conscientiousness and internal locus of control, based on data on adult respondents from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. Using data from the National Longitudinal Survey of Youth 1979 (NLSY79) Heckman et al. (2006a) find that locus of control is affected primarily by high school attendance and not by college attendance, which points towards a potential point of exhaustion of malleability in late adolescence. In an Australian sample, Schurer et al. (2015) find that tertiary education still plays a role in shaping non-cognitive skills of young adults, as they find a positive impact of university education on extraversion.

Thiel et al. (2014) exploit the same high school reform as we do – the reduction in the length of academic-track high school – in one East German federal state (Saxony-Anhalt).[18] They find some significant effects of increased learning intensity on adolescent personality development, which they, however, do not consider to be economically meaningful. Their analysis focuses, however, on only one of Germany's 16 federal states. Since this East German state prolonged high school by one year only shortly before the reform (following German reunification), the implementation and consequences of the reform may not be representative for the whole of Germany. Furthermore, since time and reform effects cannot be separated, their estimates might be biased by other factors affecting the region. Moreover, their analysis only takes into account the double cohort of graduates, that is, the last cohort graduating from high school with a total of 13 years of schooling and the first cohort graduating after 12 years. This double cohort of graduates undoubtedly features peculiarities that may lead to confounding effects and potentially offset true effects of the reform.[19]

Our study is unique in the sense that it uses a nationwide quasi-experiment and representative data to investigate the causal impact of education on personality. Furthermore, none of the previous studies investigated heterogeneous effects, distinguishing between subgroups of students that may have been affected differently by the educational reform. Finally, our rich information on individual characteristics and activities allow us to investigate potential mechanisms by which the reform affected personality traits.

[18] For a detailed description of the high school reform and the German education system in general, see Section 2.3.

[19] For example, increased competition due to a limited number of available jobs may lead students who graduate within 13 years to be worried or feel pressure because employers may prefer job candidates from the younger cohort given that they hold the same educational degree. As a consequence, any potential increase in stress among the younger cohort as a true effect of the reform, for example due to a higher workload, would therefore be offset and would not be measurable.

2.3 The German High School Reform

2.3.1 Institutional Background

In Germany, educational policy is the responsibility of the federal states (*Bundesländer*). In all of the states, children start elementary school at the age of six and usually continue to secondary school after four years.[20] The German educational system comprises three basic types of secondary school: the lower, *Hauptschule*, the intermediate, *Realschule*, and the higher track, *Gymnasium*. Only successful completion of the *Gymnasium* (henceforth referred to as academic-track or simply high school) leads to the university entrance qualification (*Abitur*). Some states also have comprehensive schools, which combine the aforementioned three secondary school types.

Until 2001, high school lasted nine years in almost all federal states (except for Saxony and Thuringia), resulting in a total of 13 years of schooling to obtain a high school diploma.[21] The relatively long time spent in the university preparatory track compared to most other industrialized countries was considered a disadvantage for German graduates, who therefore entered the labor market later than in other countries. Starting in 2001, most federal states introduced a school reform to shorten the length of high school from nine to eight years for newly entering students, reducing overall schooling from 13 to 12 years.[22] This reform allowed high school students earlier graduation and earlier labor market entry, which was attractive for several reasons.

First, the reform envisaged that graduating younger would help German school-leavers to be more competitive in the international labor market. Second, by shortening secondary school duration, the reform was to make the German education system less costly per student and hence more efficient. Finally, the reform was meant to address the challenges emerging in many Western countries due to demographic changes. By expanding the labor force by one birth cohort, the reform was designed not only to reduce the shortage of skilled workers but also to solve the problem of an increasing disparity between a reduced group of young workers contributing to the pension scheme and a rising share of an older population receiving pension benefits. In contrast, opponents of the reform feared that shortening the high school track

[20] An exception are the states Berlin, Brandenburg and Mecklenburg-West Pomerania, where the placement to different secondary school tracks takes place at grade seven, as opposed to grade five.

[21] In the former East, students finished secondary school after 12 years. Following reunification, the East German states of Brandenburg, Mecklenburg-West Pomerania, and Saxony-Anhalt adapted to West German standards and increased total years of schooling from 12 to 13 in the 1990s. In contrast, Saxony and Thuringia have kept 12 years of schooling.

[22] In the states of Berlin, Brandenburg, and Mecklenburg-West Pomerania, where tracking takes place in grade seven, the reform reduced high school from seven to six years. In the following, whenever we refer to the year of high school entry, we mean the year in which students began grade five and pursued the academic track. Analogously, high school refers to the last eight (or nine) years of school from grade five until graduation.

would harm the quality of education.

The German high school reform was introduced in most of the federal states between 2001 and 2007.[23] Exceptions are Rhineland-Palatinate, where the length of high school has been reduced only in selected schools so far, and Saxony and Thuringia, which have not reduced the length of high school. The educational reform was first implemented in the state of Saarland in the 2001/2002 school year. Other states followed up to 2007. Since the reform was mandatory for all high schools within a state, it was almost impossible for students to avoid the change in the educational system.[24] A detailed overview of the introduction of the reform and the first cohorts to be affected, broken down by federal state, is given in Table 2.1.

Table 2.1: Introduction of the Reform by Federal State

Federal State	Implementation of the reform	Graduation of first cohort affected
Saxony[a]	—	—
Thuringia[a]	—	—
Saarland	2001	2009
Hamburg	2002	2010
Saxony-Anhalt	2003	2007
Mecklenburg-West Pomerania	2004	2008
Bavaria	2004	2011
Lower Saxony	2004	2011
Baden-Wuerttemberg	2004	2012
Bremen	2004	2012
Hesse[b]	2004	2012-2014
North Rhine-Westphalia	2005	2013
Berlin	2006	2012
Brandenburg	2006	2012
Schleswig-Holstein	2007	2016
Rhineland-Palatinate[c]	2007	—

[a]Saxony and Thuringia kept the 12-year school system after reunification.
[b]Gradual introduction for students entering high school in 2004/05 (10% of all schools), 2005/06 (60%), and 2006/07 (30%).
[c]In Rhineland-Palatinate, the reform has only been introduced in selected schools so far.
Source: Autorengruppe Bildungsberichterstattung (2010)

While it did reduce the total duration of schooling, the reform did *not* reduce the overall curriculum. The total number of hours required for graduation (265 year-week hours) between grade five and graduation were maintained (see KMK, 2013). As a result, students attended on average more than 33 hours per week in school after the reform, compared to less than 30 before the reform.[25] The decision on how to distribute the total year-week hours over the remaining

[23]Current initiatives in some states to return to 13 years of schooling are only debated since recently, and, therefore, do not concern our sample of students described in Section 2.4.
[24]See Section 2.6.5 for a discussion on non-compliance with the reform.
[25]Note that an increase in class hours is accompanied by an increase in homework. Therefore, the true increase

grades was left up to the individual schools, but the great majority of schools allocated the increase in hours to grades seven to nine, when students are aged 13 to 16, such that at these ages students attended even up to 36 hours per week. This increase in workload per unit of time led to a higher learning intensity (per week)[26], and prolonged school days.[27]

2.3.2 Anticipated Effects of the Reform on Personality

Since the overall curriculum and the requirements for the German high school diploma were left unchanged, the policy reform increased the learning intensity for students through a greater number of hours per school year. This might have had various unintended consequences such as an increase in grade repetition (Huebener and Marcus, 2015) or a decrease in mental health due to a higher stress level (Quis, 2015).

A similar change in the German education system took place in the 1960s, moving the start of the academic year to an earlier date. In 1966 and 1967, the academic year was shortened, reducing the total amount of time in school by two-thirds of a year for students enrolled in school at that time, while the basic curriculum was left unchanged. Pischke (2007) finds that there were no adverse effects of shorter school duration on earnings and employment later in life, and concludes that the more recent high school reform (the reform used in this study), which reduced the length of high school by one year, may not compromise labor market success of affected students in general. However, according to Pischke (2007) the shorter school years increased grade repetition in elementary schools and led to fewer students attending higher secondary school tracks. This may point to important heterogeneous effects, as the most poorly performing students appear to be less likely to keep up with the increased learning intensity. Furthermore, Pischke (2007) notes that the shorter instructional time may induce costs by shifting students' time away from activities that are not directly linked to labor market outcomes such as voting or artistic pursuits.

Since the recent reduction in the length of high school by one year represented a much greater change than the shortened academic years of the 1960s, it may well be that it also impacted students' personalities. In this study, we focus on the personality concept of the Five Factor Model (McCrae and Costa, 1999), which comprises five psychological dimensions, also

in workload per week is even higher than the pure increase in class hours.

[26] Note that some states minimally reduced the content of the curriculum by diminishing class hours of elective courses. Hence, the learning intensity did not increase as much as it could have, and we estimate lower bound effects of the reform.

[27] At the same time, there was a substantial increase in all-day high schools. By 2010, almost half of all high schools had become all-day schools compared to only about 12% in 2002 (Autorengruppe Bildungsberichterstattung, 2012). In addition to the compulsory curriculum, most of these all-day schools offer further activities on a voluntary basis, which is, however, also the case for more traditional schools.

known as the *Big Five* – openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Openness to experience describes an individual's creativity and imagination, while conscientiousness refers to the propensity to work effectively, efficiently, and thoroughly as opposed to being disorganized and lazy. Extraversion is defined as an individual's tendency to be outgoing, sociable, and communicative rather than being reserved. Agreeableness describes individuals who are polite, forgiving, and kind to others. Neuroticism measures emotional instability, that is, whether a person tends to worry and get nervous easily rather than coping well with stress and being relaxed. In addition to these Big Five personality traits, we also include the concept of locus of control, which refers to an individual's perception of control over his or her life (Rotter, 1966).

Each of these personality traits may be affected differently by the different consequences of the German high school reform (cf. Thiel et al., 2014). First, the higher workload may have increased the pressure on students, and hence decreased their emotional stability, that is, increased their neuroticism. At the same time, we hypothesize that the compression of the curriculum required a higher level of self-discipline, that is, it increased students' conscientiousness about performing well in school. The reform may have had ambiguous effects on agreeableness, since, on the one hand, more cooperation may be required to cope with the higher workload, and on the other hand, students may become more selfish in a more competitive environment. Likewise, the effect on locus of control may be positive or negative depending on whether students recognize that they are able to cope or not with the increased pressure induced by external factors. Second, spending more time in school and having less leisure-time may affect social skills, that is, the agreeableness, extraversion, and openness of students. Third, the exogenous enactment of the reform may have decreased students' perceived control over their lives, that is, it may have resulted in a lower locus of control. A detailed overview of these theoretical considerations with the potential mechanisms at work is given in Table 2.2. Moreover, we expect to find heterogeneous effects for students from different socio-economic backgrounds. For example, for students from non-intact families, longer school days following the reform may have created a more stable environment due to more frequent interactions with teachers, allowing the young people to build relationships to additional adults who they can turn to for advice and support. On the other hand, these students may receive less support at home to cope with the increased workload and may therefore experience higher stress levels due to the shortening and increased intensity of high school.

It remains an empirical question which personality traits are affected by the educational change and whether or not the reform's effects vary by student characteristics.

Table 2.2: Potential Mechanisms and Anticipated Effects

Consequences of the reform and potential mechanisms	Anticipated effects
Higher requirements/ workload/ learning intensity	
Increase in pressure and fear of failure$^{(*)}$	N (+), LoC (–)
Increase in self-discipline required	C (+)
Increase in cooperation or selfish behavior	A (+/–)
Recognizing ability of coping/not coping$^{(*)}$	LoC (+/–)
More time spent in school	
Increase in social interaction with students and teachers$^{(*)}$	A (+), E (+)
Less leisure-time	
Reduction in leisure-time activities	O (–)
Reduction in time for social interaction outside school	A (–), E (–)
Exogenous enactment of reform	LoC (–)

Notes: Compare to Büttner et al. (2011); extensions denoted by $^{(*)}$.
+ (–) denotes an anticipated increase (decrease) in terms of the score of the personality traits openness (O), conscientiousness (C), extraversion (E), agreeableness (A), neuroticism (N) and locus of control (LoC).

2.4 Data and Methods

2.4.1 Sample of Adolescents and Young Adults

Our analysis is based on data from the German Socio-Economic Panel (SOEP) study, which is a representative household panel survey (Wagner et al., 2007) with around 30,000 individuals in almost 15,000 households in the 2013 wave.[28] In addition to rich information on family background and childhood environment characteristics, the SOEP provides self-ratings of personality traits for 17-year-old adolescents and for adult respondents aged 18 and over in various waves since 2005. Hence, we use data from the years 2005 to 2013, and select all adolescents and young adult respondents up to the age of 21 who were attending high school at the time of the survey or had earned a high school diploma.[29] To identify whether an individual was affected by the high school reform, we use information on the year of school entry and on the state of residence or, in cases where school has been already completed, the state where the high school diploma was obtained. In case of missings, the year of school entry was imputed from the date of birth.[30] We assign students from Saxony and Thuringia to be affected by

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

[29]Students in the academic track at a comprehensive school are affected by the reform as well. However, we exclude these students since we cannot unambiguously identify which tracks they have been in since entering secondary school. Furthermore, these students' exposure to the reform would be difficult to identify since a transitional period has been provided for implementation of the reform and exceptions are allowed at comprehensive schools in some states. Compulsory schooling ends in Germany at age 16.

[30]The year of school entry is available for 37 percent of the full sample. For these students, the imputed year matched the actual year in 93.7 percent of all cases, and the assignment to treatment or control group was correct in 99.8 percent of all cases.

the reform, even though no educational change took place, as they kept the eight-year high school track after reunification. We drop students from Rhineland-Palatinate from our sample because there the reform has been introduced only in selected schools to date. We also exclude individuals who we know repeated one or more grades to avoid noise from different levels of schooling experienced so far and from different learning intensities students were exposed to.[31]

Finally, we only include individuals who successfully answered the items from all Big Five dimensions[32] and provided valid information on family background and home environment. The final sample consists of 1,316 individuals, of whom 446 were affected by the reform.[33] A detailed description of how it is constructed is given in Table A.1 in the appendix.

2.4.2 Outcome Measures and Background Variables

The SOEP provides self-ratings of personality traits in various waves since 2005 (Gerlitz and Schupp, 2005). These cover personality measures related to the Five Factor Model (McCrae and Costa, 1999), which comprises the dimensions of the *Big Five* – openness to experience, conscientiousness, extraversion, agreeableness and neuroticism. Each of these dimensions is measured with three items, which are answered on seven-point Likert-type scales. As a further personality concept we follow the approach by Specht et al. (2013) and construct a measure for internal locus of control from seven items, which reveal whether a person believes that events are the result of his or her own actions rather than of external factors. A detailed list of all personality items is provided in Tables A.2 and A.3 in the appendix. To construct measures for the respective personality trait, the items are averaged for each dimension and standardized separately by gender at mean zero and variance one. While for adolescents these personality measures are included in the Youth Questionnaire in every year between 2006 and 2013, adult respondents' Big Five personality traits were surveyed only in the waves 2005, 2009, and 2013 and their locus of control in the waves 2005 and 2010. If an individual's personality has been measured twice, we use the personality measures at the youngest age possible, between 17 and 21 years, to guarantee a largely homogeneous sample and to avoid biased estimates through possible age effects.

To account for potential age effects we include age and age squared at the point in time of measurement in all estimations. Additionally, we control for several pre-reform individual

[31] Huebener and Marcus (2015) show that class repetitions did not increase up to grade nine following the German high school reform. Hence, the composition of our sample of 17-years-olds, who are in grade ten or eleven, should largely be unaffected by any change in the probability of repeating a grade. In our sample 68 students drop because they repeated one grade (28 from the control group and 40 from the treatment group).

[32] These are more than 94 percent of all individuals in our sample of consideration.

[33] The average size of a birth cohort surveyed at age 17 is 267 in the years under observation; on average 42 percent of the individuals in a cohort are in the high school track.

characteristics in some specifications. These include demographic and socio-economic variables, such as gender, migration background, and whether the student has lived primarily in a rural area during childhood. We incorporate students' previous educational performance, i.e. whether the student's elementary school teacher explicitly recommended not to follow onto high school, but to the lower or intermediate track instead. Variables related to family circumstances capture whether students come from a non-intact family, that is, whether they lived with only one parent up to the age of 15, the number of siblings, and birth order. Furthermore, the model accounts for parental characteristics including education, the occupational status of the father, and the employment status of the mother at the time when the individual was ten years old.[34] For a description of all variables, see Table A.4 in the appendix. The summary statistics of the personality traits and the individual characteristics are shown in Tables A.5 and A.6 in the appendix.

2.4.3 Empirical Strategy and Treatment

We exploit the German high school reform introduced in almost all federal states in the years 2001 to 2007 as a quasi-natural experiment to establish a causal effect of education on personality. We define the control group as students who entered high school prior to the reform and hence graduated after nine years of high school. Our treatment group, in contrast, consists of students who entered high school after the implementation of the reform and hence obtain their diploma after only eight years of high school. Individuals are assigned to the treatment and the control group based on their year of school entry and state of residence.[35] We exploit the variation in time and region to isolate the effect of the reform from other influential factors. To this end, we estimate the following model

$$y_{ist,17} = \alpha \text{REFORM}_{st} + X_i\beta + \sum_s (\gamma_s \text{STATE}_s + \delta_s \text{STATE}_s * t) + \sum_t \tau_t \text{YEAR}_t + \varepsilon_{ist}, \quad (2.2)$$

where $y_{ist,17}$ is the personality measure at age 17 (or slightly older, as described in Section 2.4.1) of person i in state s who has entered high school in year t. Our prime variable of interest, $REFORM_{st}$, equals 1 if in state s students entering high school in year t are affected by the reform, and 0 otherwise. $STATE_s$ is a set of state dummies, $YEAR_t$ dummies indicating the year of high school entry, and $STATE_s * t$ state-specific linear trends in these years. X_i

[34] In an additional estimation, we include parental personality traits. These may be problematic as for the majority of the sample, parental personality is not observed pre-reform and, hence, may entail spill-over effects from children to their parents. Therefore, we exclude parental personality traits in our preferred specification. Still, our estimates are robust.

[35] Individuals who already graduated from high school are assigned to the treatment or control group based on the state where their high school diploma was obtained.

is a vector of individual characteristics, including age and age squared, gender, and, in some specifications, a number of pre-reform characteristics, which may be correlated with personality. These include the occupational status of the father, parental education, childhood environment, a dummy for non-intact family, and the employment status of the mother when the individual was ten years old. Furthermore, we include indicator dummies for the different SOEP subsamples and we cluster error terms at the state level.[36] When estimating equation (2.2), our main parameter of interest is α, which indicates the impact of the high school reform on the respective personality trait. Our analysis therefore resembles an extended difference-in-differences setting, in which the key identifying assumption is that, in the absence of a reform, α will be zero, that is, the treated ($REFORM_{st}$=1) and the non-treated ($REFORM_{st}$=0) do not differ significantly in terms of their personality traits.

Next, we turn to the description of the exact treatment that we are able to measure in our analysis. Although the treatment group experiences a shortened high school track compared to the control group, we do not measure the effect of a decrease in years of overall school duration per se, since the majority of our sample was surveyed the year they turned 17 and in most cases were still enrolled in high school. Instead, the focus of our study is on the compression of learning achievement and how this shapes personality traits during adolescence. Hence, the treatment implied by the reform is a higher workload in school, especially between ages of 13 and 16. Affected students therefore experienced a higher learning intensity during these years, spent more time per day in school, and had a higher level of learning achievement by age 17 because of more cumulative hours of school.

2.4.4 Timing of the High School Reform and Common Trend Assumptions

The key identifying assumption of our analysis is that there are no other policy changes or regional shocks that coincide with the implementation of the high school reform and affect students' personality traits. Identification, therefore, requires that the outcomes of interest, that is, the personality traits, of students affected by the reform and students not affected by the reform would have followed the same trend over time had the high school reform not been implemented. This *common trend assumption* boils down in our case to the assumption that the personality of high school students would have evolved similarly in the different federal

[36] To account for the small number of clusters, it may be necessary to use wild cluster bootstrapped standard errors (see Cameron et al., 2008). Our estimations, however, show that the wild cluster bootstrap rather leads to even slightly lower standard errors if different at all. We therefore report the usual standard errors without bootstrapping, as it is the more conservative estimation method in our case.

states. While this is not testable, as the counterfactual observation in absence of the reform cannot be observed, we argue that the assumption is realistic and nonrestrictive. We exploit the variation in the introduction of the reform over time, as we compare high school students from different federal states in Germany and control for a potential time trend through year fixed effects. We believe this is the least restrictive possible comparison as the evolution of one student's personality is likely to most closely resemble that of another student in the same school track (high school) even if the two are from different states (as opposed to students from the same state but different school tracks). As we control for state fixed effects, students from different federal states are allowed to differ in their levels of personality traits. Further, we include state-specific linear time trends, such that even slopes of personality development over time may differ across states.

Moreover, crucial for the causal interpretation of the reform effect is that the variation in schooling is truly exogenous to serve as a quasi-natural experiment. This assumption would be violated if there were self-selection into either the treatment or the control group or if the assignment were not random. In this analysis, neither should be an issue. Since the reform was introduced in an entire state at one time, the only possibility to avoid the reform, that is, to self-select into the control group, would have been to move to a different state.[37] Given the high costs associated with moving a whole family to another state, this seems highly unlikely.[38] This is also confirmed by the descriptive statistics on individual characteristics of the control group and the treatment group, which are provided in Table A.6 in the appendix. Mean comparisons of the treatment and control groups show that no (pre-reform) variables exhibit severe differences on average, apart from *age* and *East*. These differences can be explained by the composition of the sample and the earlier introduction of the reform in the East German states.

Different dates of implementation of the reform were chosen to avoid overcrowding of universities, and the timing of the high school reform may be related to specific state characteristics. Since we control for state fixed effects in our analysis, it is not necessary that the timing of state reforms be unrelated to state characteristics (see Black et al., 2005). Nevertheless, we investigate whether the timing of the implementation of the new education system followed some pattern. For this purpose, we ran OLS and ordered probit regressions of the timing of state school reforms (late implementation, year in which the reform was implemented) on various pre-reform state characteristics (see Table A.7 in the appendix). The analysis shows that there is no such pattern, as the timing of the reform is not related to the percentage of high school

[37] The only exception to this is Hesse, where there are double graduating cohorts in three consecutive years, 2012, 2013, and 2014. We therefore only include students from Hesse who were not affected by the reform and who graduated in 2012 or earlier and students who graduate in 2014 or later and were affected by the reform.

[38] See Section 2.6.5 for a more extensive discussion of selectivity and robustness checks.

students in a state's population, to whether the state government is conservative, to whether the next state elections were scheduled for 2001/2002, or to the state's GDP per capita. There is some evidence that states with a higher median age of residents adopted the high school reform slightly earlier, which is an artifact caused by the older population in East German states.

2.5 Results

The results of our OLS estimates of equation (2.2) are provided in Table 2.3.[39] The first specification does not control for any individual characteristics apart from gender, age and age squared to account for any potential age effects. The second specification additionally controls for several individual pre-reform characteristics. The parameter estimates of the reform effect are almost identical in both specifications. This manifests the validation of using the high school reform as a quasi-natural experiment, as individual characteristics which may be correlated with personality do not appear to be correlated with the reform, and their omission does not bias the estimated impact. Therefore, in our main specification we only control for this very parsimonious set of individual characteristics, next to fixed effects for state, year of school entry, SOEP sample, and state specific linear time trends.

There is no clear impact of the reform on openness, conscientiousness, and extraversion in the overall population of high school students as coefficients are small in size and statistically insignificant. In contrast, the estimates show that the reform significantly increased students' neuroticism. The effect is substantial in that emotional stability is altered by more than a third of a standard deviation on average. Compared to the predictive power of adverse childhood conditions on adult personality, the effect size is substantial: an increase of 0.37 standard deviations in neuroticism due to the reform, exceeds that of ADHD experience (0.22 standard deviations) and is only slightly lower than that of sexual abuse during childhood (0.41 standard deviations) as investigated by Fletcher and Schurer (2015) in a sibling-fixed effects estimation. This increase in neuroticism may reflect the impact of the higher workload, which may have increased stress and pressure on all students. Following the reform, agreeableness and locus of control reveal changes of 0.12 to 0.14 standard deviations. These effects are, however, not statistically significant.[40] Yet, their directions are in line with the expectation that changes in

[39] Table 3 provides results from unweighted regressions. However, weighted regressions deliver virtually the same estimated coefficients.

[40] Note in general, that due to the relatively small number of observations, we pick the 10% level of significance to be relevant. Given this small sample size, the fact that some of the coefficients appear to be statistically significant at this, or even more conservative levels, makes our findings even stronger. For the interpretation of any other coefficient, one should keep in mind that statistical insignificance does not prove that there is no

Table 2.3: Effects of the Reform

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Reform	-0.034	-0.088	0.035	0.135	0.370**	-0.122
	(0.148)	(0.178)	(0.126)	(0.119)	(0.150)	(0.108)
Female	0.021	-0.017	-0.014	-0.010	0.006	-0.013
	(0.062)	(0.073)	(0.064)	(0.054)	(0.058)	(0.070)
Age	-0.676	0.920	0.238	0.406	1.933**	-0.661
	(1.344)	(1.130)	(1.206)	(0.863)	(0.843)	(1.338)
Age^2	0.017	-0.022	-0.006	-0.011	-0.051**	0.015
	(0.036)	(0.030)	(0.032)	(0.023)	(0.022)	(0.035)
R^2	0.037	0.053	0.049	0.044	0.068	0.066
Observations	1316	1316	1316	1316	1316	1202
Reform	-0.033	-0.102	0.035	0.135	0.368**	-0.112
	(0.151)	(0.167)	(0.119)	(0.117)	(0.150)	(0.112)
Female	0.024	-0.023	-0.017	-0.006	-0.009	-0.006
	(0.062)	(0.070)	(0.060)	(0.057)	(0.059)	(0.068)
Age	-0.831	1.283	0.209	0.442	1.993**	-0.761
	(1.258)	(1.255)	(1.207)	(0.918)	(0.848)	(1.356)
Age^2	0.021	-0.032	-0.005	-0.012	-0.053**	0.018
	(0.033)	(0.033)	(0.032)	(0.024)	(0.022)	(0.036)
Rural area	-0.067	-0.042	0.021	-0.010	0.103**	-0.026
	(0.064)	(0.100)	(0.075)	(0.061)	(0.043)	(0.055)
Non-intact family	0.148***	-0.099	0.172*	-0.050	0.039	-0.119*
	(0.040)	(0.081)	(0.083)	(0.057)	(0.081)	(0.063)
Working-class father	-0.268***	0.044	0.002	-0.055	0.117	0.008
	(0.086)	(0.075)	(0.079)	(0.078)	(0.091)	(0.066)
High parental educ.	0.040	-0.199***	-0.127	-0.029	0.014	0.049
	(0.058)	(0.044)	(0.073)	(0.058)	(0.041)	(0.053)
Working mother	0.071	-0.057	0.129**	-0.033	-0.150***	-0.013
	(0.066)	(0.063)	(0.055)	(0.042)	(0.044)	(0.062)
Migration background	0.226**	-0.033	-0.066	-0.061	0.005	0.012
	(0.090)	(0.082)	(0.073)	(0.086)	(0.094)	(0.058)
Low-perform. student	-0.067	-0.068	-0.277**	0.084	0.092*	-0.188
	(0.121)	(0.088)	(0.113)	(0.117)	(0.049)	(0.110)
R^2	0.057	0.064	0.066	0.046	0.077	0.072
Observations	1316	1316	1316	1316	1316	1202

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. State specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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.

the type and intensity of social interactions between students due to longer school days may be beneficial for agreeableness, while the exogenous enactment of the reform may decrease the

effect, but only that we fail to reject it, which may be an artifact of the sample size.

perceived control over one's life. Considered together, these findings suggest that the change in the educational system may impact particular facets of personality, possibly via different channels.[41]

With respect to the individual characteristics, we find only few meaningful driving forces, especially among parental characteristics.[42] Living in a non-intact family appears to be related to several personality traits. We observe a higher level of openness and extraversion and a lower level of perceived control. The former effects could manifest children living with only one parent being more familiar to new and changing environments and, therefore, more open and extroverted towards new experiences and others. The latter effect on locus of control also is in line with general expectations considering children of separated parents feel that they can do little to change the status quo. Yet, they do not seem to suffer from significantly more emotional stress as they do not differ substantially in their level of neuroticism from other students. The more alarming it even is that the change in the educational system has a larger impact on neuroticism than does a disrupted family, both in terms of magnitude and significance: the high school reform increases neuroticism by almost ten times as much as being raised in a non-intact family.[43]

Moreover, students with a lower socio-economic background, that is, students from working-class or less-educated families, are on average less open and more conscientious than students with higher socio-economic backgrounds.[44] On average, less-educated parents with blue-collar jobs tend to be characterized by hard work and social conformity rather than creativity, which may be associated with less openness and more conscientiousness. Children might either inherit these traits from their parents or be socialized in a particular way, for example, by learning from their parents as role models throughout life. Students with a working mother are on average significantly more extroverted and less neurotic than students whose mother did not work when they were ten years old. This could be due to the fact that children of working mothers learn independence and social interaction relatively early through increased exposure to day care. They may therefore be more outgoing and settled in life, leading to higher extraversion and emotional stability.

Students with a migration background display a higher level of openness than students

[41] For a more extensive discussion on the potential mechanisms at work, see Section 2.5.2.

[42] Note that due to the standardization separately by gender, all differences in absolute levels of the scores on personality traits between male and female students are erased. We therefore expect the coefficient of *female* to be small and insignificant in the estimation, which is indeed the case. However, this does therefore not imply that there are no significant differences by gender in the original scores on personality.

[43] Note that this applies to the sample of high school students. Being raised in a non-intact family may have larger effects on neuroticism in the overall population.

[44] Note that our sample is special in the sense that it only comprises high school students. As these students are in a presumably more advantaged situation than students enrolled in other school tracks, the effects of the individual characteristics on personality may not be representative for the average adolescent in Germany.

without a migration background, but do not exhibit further significant differences in traits. Students growing up mainly in a rural area show greater levels of neuroticism, as well as low-performing students do. In addition, low-performing students have significantly lower levels of conscientiousness which can explain their lower achievements in the past.

In an extended specification, we further include family characteristics related to the number of siblings and birth order for a subsample of students with the relevant information (see Table A.8 in the appendix). While the reform effects prove to be stable, these estimates reveal that first-born individuals are more conscientious. In an additional specification, we include the respective parental personality traits in each regression for a subsample of our students (see Table A.9 in the appendix). The respective coefficients stand out in both magnitude and significance, pointing to a strong intergenerational transmission of personality.[45] However, their inclusion makes little difference concerning the reform coefficients.

2.5.1 Heterogeneous Effects

So far, the estimates show average effects for the overall population of high school students. Following the high school reform, students tend to be less emotionally stable, while some students may also be slightly more agreeable but perceive to have less control over their life. To truly shed light on the effects of the reform, however, it is important to consider how the policy change has affected particular subgroups of students differently. This may not only reveal which students' personality traits are especially susceptible to a change in the educational system, but possibly also illustrate the different mechanisms by which the reform impacts personality. A natural distinction here is by *gender*, since boys and girls of a given age not only differ in their stage of physical and mental development but also in their behavior. As shown in Table 2.4, the decrease in locus of control is driven by female students, for whom this effect is statistically significant and economically large, amounting to one-fourth of a standard deviation in magnitude. In contrast, the increase in agreeableness remains statistically insignificant for both groups. Still, it is mainly prevalent among male students only. Since boys are usually observed to be more competitive and less agreeable than girls, the exogenous variation in social interaction or necessity for cooperation may have forced boys in particular to become more sociable and cooperative. This finding therefore illustrates how a change in the educational system also could mitigate gender differences in personality or behavior and bring initially different levels of development into closer alignment across genders.

[45] The intergenerational effects are between 0.082 and 0.261 for the Big Five personality traits and 0.177 and 0.141 for locus of control. These results are largely in line with the intergenerational correlation coefficients reported by Anger (2012) for children of all school types.

Table 2.4: Heterogeneous Effects of the Reform

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Female						
Reform	0.015	-0.039	0.047	0.196	0.342*	-0.008
	(0.152)	(0.160)	(0.126)	(0.142)	(0.176)	(0.110)
Interaction	-0.096	-0.098	-0.023	-0.120	0.054	-0.235**
	(0.116)	(0.122)	(0.136)	(0.137)	(0.092)	(0.094)
East						
Reform	0.074	-0.058	0.080	0.196*	0.251	-0.134
	(0.157)	(0.176)	(0.165)	(0.106)	(0.167)	(0.150)
Interaction	-0.423	-0.128	-0.162	-0.244*	0.468*	0.063
	(0.326)	(0.320)	(0.288)	(0.121)	(0.228)	(0.248)
Non-intact family						
Reform	-0.129	-0.055	-0.088	0.131	0.403**	-0.100
	(0.151)	(0.159)	(0.135)	(0.127)	(0.153)	(0.126)
Interaction	0.388***	-0.124	0.505***	0.030	-0.153	-0.072
	(0.072)	(0.156)	(0.132)	(0.086)	(0.156)	(0.160)
Working-class father						
Reform	-0.012	-0.093	0.004	0.173	0.407**	-0.095
	(0.139)	(0.175)	(0.121)	(0.123)	(0.150)	(0.097)
Interaction	-0.060	0.001	0.160	-0.192	-0.229	-0.134
	(0.162)	(0.137)	(0.142)	(0.153)	(0.146)	(0.206)
High parental education						
Reform	-0.031	-0.085	0.119	-0.001	0.345**	-0.221
	(0.211)	(0.135)	(0.198)	(0.127)	(0.138)	(0.175)
Interaction	0.004	-0.025	-0.154	0.233**	0.040	0.172
	(0.144)	(0.107)	(0.154)	(0.095)	(0.045)	(0.138)
Working mother						
Reform	-0.126	-0.245	0.037	0.124	0.380	-0.154
	(0.147)	(0.156)	(0.161)	(0.153)	(0.222)	(0.167)
Interaction	0.135	0.218	0.006	0.015	-0.023	0.043
	(0.152)	(0.190)	(0.150)	(0.156)	(0.136)	(0.159)
Migration background						
Reform	-0.110	-0.161	0.032	0.148	0.356**	-0.078
	(0.122)	(0.187)	(0.102)	(0.142)	(0.154)	(0.118)
Interaction	0.455**	0.437**	0.019	-0.074	0.082	-0.295
	(0.162)	(0.167)	(0.219)	(0.221)	(0.187)	(0.284)

Continued on next page

Continued from previous page

Low-performing student						
Reform	-0.020	-0.078	0.039	0.141	0.364**	-0.136
	(0.146)	(0.172)	(0.122)	(0.119)	(0.148)	(0.112)
Interaction	-0.143	-0.113	0.022	-0.079	0.039	0.195
	(0.196)	(0.246)	(0.346)	(0.143)	(0.150)	(0.201)
Observations	1316	1316	1316	1316	1316	1202

Notes: SOEPv30 waves 2005 to 2013. OLS regressions, separately for each interaction considered. Only the coefficients of *REFORM* and the respective interaction with *REFORM* are presented. Female, age, age squared, the respective control variable of interest (without interaction with reform), state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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.

Additionally, we investigate differential impacts of the high school reform in former *East* and *West Germany*, as there are not only persistent differences with respect to the socio-economic environment, but also differences in educational policies between the two parts of Germany. Especially the long-standing tradition of the former 13-year school system is not present in the former East. Those Eastern states that reduced high school by one year in the 2000s prolonged high school by one year just a decade before. This could mitigate any changes induced by the recent high school reform if fewer infrastructural adjustments were necessary for the switch back, compared to former West Germany. In line with these expectations, the estimates in Table 2.4 reveal that the increase in agreeableness is entirely driven by students in West Germany. Contrary to this intuition, however, the increase in neuroticism is significantly more pronounced among students in East Germany. One reasonable explanation is that in Eastern states, where students normally attend elementary school for their first six years of schooling, seventh-grade students were confronted with the transition to high school and with a higher workload at the same time. In contrast, West German seventh-graders already had two years to acclimatize themselves to high school before being confronted with the higher workload. Hence, this finding may demonstrate the importance of gradually changing the learning environment of students who may feel stressed if several features are altered at the same time.

In the next step, we look into heterogeneous effects by family structure. For students from *non-intact families*, the increase in time spent at school may have created a more stable environment with stronger interactions with teachers as additional adult reference persons. In contrast, these students may also receive less support at home to cope with the increased learning. The estimates provide suggestive evidence for the former hypothesis; students who did not live with both parents during their entire childhood benefit from the educational change in terms of openness and extraversion. In contrast, there is no evidence to support the hypothesis

that these students have an even higher increase in neuroticism due to a potential lack of support at home. We hypothesized that children with a large *number of siblings* and with a high *birth order* may have greater difficulties coping with the higher workload after the reform as they may not receive the same amount of parental support as a first child or only child. We find suggestive evidence only for the latter hypothesis as first-borns face a significantly lower increase in neuroticism than students of higher birth order.[46]

Furthermore, it is of interest how students from different socio-economic backgrounds respond to educational changes. To this end, we investigate heterogeneous effects by the *occupational status of the father* as well as by *parental education*. It is conceivable that students with a lower socio-economic status are less likely to receive (adequate) support from their parents to meet the increased demands at school. Hence, the higher workload induced by the reform may lead to a higher stress level for this group. However, we do not find any significant differences in the impact of the reform on neuroticism, neither by occupational status of the father nor by parental education. The estimates even indicate a smaller increase in emotional instability for students with a working-class father, though this difference is statistically not significant. In contrast, agreeableness increases especially among the more privileged students with better educated parents.

Moreover, we investigate whether students with a *working mother* respond differently to the high school reform. Since the reform led to longer school days and thus to a better infrastructure for organized afternoon activities, students with working mothers may benefit from having improved supervision in the afternoon. However, we do not find such differences in the impact of the reform by the employment status of the mother.

Furthermore, students with a *migration background* may respond differently to the reform than students from non-migrant families. One could hypothesize that the former benefit from longer school days by becoming better integrated into their peer groups and by improving their language skills. If so, this reform could help to reduce initial and persistent differences between these two groups and foster integration and educational performance of students from migrant families. Indeed, we find substantial differential effects of the reform on openness and conscientiousness. Both traits increase by more than 0.4 standard deviations for migrant compared to non-migrant students. Lundberg (2013) finds openness and conscientiousness to be important correlates of later educational achievement, in particular depending on socio-economic background. These reform effects may, therefore, be important in fostering equality of opportunity between individuals with and without migration background.

[46] Results are available from the authors upon request.

A general concern is that *low-performing* students are particularly vulnerable to the educational reform as they are most at risk of being 'the first to be left behind' by the higher requirements at school. Hence, we investigate whether students with low performance prior to high school show differential effects. We define those students as low-performers who, at the end of elementary school, received the teacher recommendation to follow on to the basic or the intermediate track, but not to high school. The increase in neuroticism following the reform is, however, not statistically different for this group.[47]

2.5.2 Potential Mechanisms

So far, we find evidence of a causal effect of the high school reform on adolescents' personality traits, in particular on emotional stability. The other traits changed in particular for specific subgroups. However, the effect of the reform on students' personalities still remains largely a black box because the compression of schooling had a number of implications. Affected students faced a higher learning intensity especially between ages 13 and 16, spent more time per day in school, and were expected to achieve a higher level of learning achievement by age 17 due to the higher number of cumulative school hours. Hence, the reform could have impacted students' personalities through various distinct channels.

First, the increase in learning intensity due to the compressed curriculum may have led to higher pressure on students, leading to adverse health effects. Previous research on working conditions and health has shown that an increase in workload negatively affects health related outcomes (Proctor et al., 1996), which may be related to an individual's emotional stability.[48] To shed more light on this potential mechanism, we investigate whether the reform has impacted students' health related characteristics. Indeed, we find that the reform had adverse effects on the perceived health status of the students (see Table 2.5): The likelihood to report good or very good health decreases by 5.4 percentage points. This is in line with findings on the double cohort in the state of Baden-Wuerttemberg by Quis (2015) who shows that the reform led to more stress and more mental health related symptoms, in particular among females. This could also explain why females react more delicately to the exogenous enactment of the reform in terms of their perceived control. Furthermore, we find that shortening the high school track increases feelings of sadness. Following the reform, the probability of being sad at least sometimes in the past four weeks (as opposed to seldom or very seldom) rose by 17.8 percentage points among

[47] Note, however, that the share of low-performing students, identified through the teacher's recommendation at the end of elementary school, is naturally very low among high school students, making it difficult to clearly interpret our results.

[48] Taking automotive workers as an example, Proctor et al. (1996) find overtime work to be associated with increased feelings of depression, fatigue, and confusion.

affected students. We take these results as suggestive evidence that health and emotions are a relevant transmission channel.

Table 2.5: Effects of the Reform: Mechanisms

	Outcome Variables					
	Health & Emotions		Satisfaction with School performance		Leisure-time Activities	
	Health	Sadness	School	Literature	Music	Sport
Reform	-0.054*	0.178*	-0.039	-0.089*	0.069	-0.039
	(0.026)	(0.090)	(0.058)	(0.050)	(0.063)	(0.059)
Observations	1315	768	744	744	911	910

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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.

Moreover, the high learning intensity may have directly affected students' performance. As shown by Proctor et al. (1996), an increase in workload is in some cases significantly associated with impaired performance on tests of attention and executive function. In the educational context, an increase in learning intensity may therefore also decrease attention and cognitive functioning among students, resulting in inferior school performance. Büttner and Thomsen (2015) show for the double cohort in the state of Saxony-Anhalt that an increase in learning intensity indeed negatively affected students' grades in Mathematics.[49] This in turn may have increased the feeling of pressure and stress due to fear of failure among students who were affected by the reform. In additional regressions, we actually find that the high school reform decreased students' satisfaction with their school performance. The share of students who were very satisfied with their performance in the subject German literature decreased by almost nine percentage points (see Table 2.5).[50] Unreported regressions show that this decrease is larger in East Germany. Even though this difference is not statistically significant, it is in line with the decrease in emotional stability being much more pronounced in the East. For overall school performance the drop in satisfaction is about half the size, yet not significant. In addition, we find suggestive evidence that the high school reform increased students' need

[49] Unfortunately, we cannot investigate reform effects on school performance, as the students' school grades are not comparable across our sample because there are no central exams over all states. Moreover, while the grades of students in our control group only counted on their annual school report card, the grades of students in our treatment group were already counted into their final high school grades and hence were much more relevant.

[50] We define students as being very satisfied with their educational performance if they rate their satisfaction either 9 or 10 on a scale from 0 (low) to 10 (high). Among the control group, 19 percent reported very high satisfaction with school performance overall and 23 percent with literature. Levels of satisfaction with the subjects Maths and first foreign language were not significantly affected by the reform.

for paid tutor lessons (not reported).[51] Overall, we conclude that the increase in workload following the high school reform caused lower school performance, worse health conditions, and emotional difficulties among students, which explains the decrease in emotional stability particularly among East German students.

Second, the change in the institutional framework may have affected students through the change in time allocation, as those who were affected by the reform spent more time per week in school. They not only attended additional classes, which often took place in the afternoon, but also spent lunch together, which in most schools was introduced with the reform. Hence, the reform led to more social interactions with teachers and other students and fostered independence from the parents. This stimulation of interpersonal behavior may positively affect openness, extraversion, and agreeableness, which increased especially among particular subgroups. At the same time, however, students have less leisure-time, which could adversely affect personality traits related to social behavior. Furthermore, openness could be reduced through less diversity in the activities students engage in. To examine whether and how students' time allocation was affected by the reform, we investigate their participation in two leisure-time activities: music and sports (see Table 2.5). We do not find any evidence of the reform having affected participation in either of these activities.[52] Both the treatment and the control group seem to engage equally in both. These findings suggest that our results are more likely driven by the stronger student-teacher or student-student interactions as a result of longer school days and not by the change in time allocation away from non-academic activities. This also is in line with the findings on heterogeneous effects: as mostly disadvantaged students with a non-intact family or migration background benefit in these traits, the changes induced by the reform seem to provide a more stimulating environment than the counterfactual. Yet, for all other students, stimulating counterfactual environments were not compromised.

Third, the higher level of learning achievement by age 17 due to more cumulative school hours may have impacted personality traits through changes in cognition. Since the great majority of schools increased hours between grades seven and nine (ages 13-16), the adolescents in our treatment group have typically accumulated a full year of additional learning over the control group. The earlier learning may increase cognitive skills at a younger age. Therefore, Andrietti (2015) and Dahmann (2015) investigate students still attending high school, aged 15 and 17, respectively. While effect differences with respect to the dimension of cognition and

[51] Following the reform, the percentage of students attending paid tutor lessons in addition to their regular school attendance increased by six percentage points, though we fail to prove statistical significance at conventional levels.

[52] This holds true not only for *whether* students participate in these activities at all, but also for the *frequency* of participating in these activities.

gender are observed, both studies find cognitive skill measures to increase as a result of the reform. This in turn may improve non-cognitive skills through the dynamic complementarity and the cross-effects suggested by Cunha and Heckman (2007) in their skill formation model.

2.5.3 Cost-Benefit Calculations

Ideally, the changes in personality identified in Section 2.5 could be used for an evaluation of costs and benefits induced by the high school reform. However, this is particularly difficult for two reasons. First, the lack of reliable data on the increase or reduction in costs due to the policy change, and second, because it requires the assumption that effects are quantifiable. The latter is particularly critical when studying intangible effects like those on personality. Still, when quantifying the results through prospective earnings, which is only one possibility,[53] a simple back-of-the-envelope calculation can give an idea of how to put the identified effects into perspective.

Table 2.6 presents a simplified overview over possible wage costs and benefits induced by the changes in personality traits. Since the analyses show important heterogeneous effects of the reform on personality, we consider examples of different groups: average male and female students, students from non-intact families, and students with migration background. Average male students showed only significant increases in neuroticism, while female students showed, additionally, a decrease in perceived control. Students from non-intact and migrant families turned more open and extroverted, and more open and conscientious, respectively, in addition. We compute wage returns to these changes based on estimates for Germany by Heineck and Anger (2010)[54] and for the Netherlands by Nyhus and Pons (2005)[55]. These two studies also reveal that wage returns to personality may severely differ across gender: while openness is for example penalized among male employees, female employees are rewarded for higher levels of openness. For this reason, we distinguish students from non-intact and migrant families further by gender as well.[56] Under the assumption that the changes induced in personality

[53] Certain personality traits may be beneficial for one dimension, like the labor market, while detrimental for another, like social life or health. Therefore, focusing for example on prospective earnings as one aspect only, naturally renders an incomplete description of overall costs and benefits, and alternatives should be considered in addition.

[54] We take significant coefficients of openness, conscientiousness, extraversion, and locus of control from OLS regressions as the most conservative estimates. We reverse their coefficient on external locus of control to apply it to our measure of internal locus of control.

[55] We rely on estimates on emotional stability as the reverse of neuroticism from this study. The analysis by Heineck and Anger (2010) includes effects of neuroticism as well. However, in contrast to previous literature they do not find significant effects arguing that this may be due to controlling for an individual's attitude towards reciprocal behavior, as both are linked. We, therefore, prefer to use estimates from a study without controlling for reciprocity as in the present study it cannot be accounted for as a potential outcome either.

[56] We cannot do this for the actual reform effects as this would require a model where not only e.g. migration status but also gender is interacted with the reform and with each other, which is unfeasible to interpret with the current sample size. The distinction is, however, important for multiplying the personality changes with the

traits persist into adult working ages, we find that the reform in all groups led to decreases in wages[57] (column 1). However, an important benefit of the reform is that it enables individuals to enter the labor market one year earlier, thus expanding total working years by one and, hence, increasing lifetime earnings. For this reason, we present changes in hypothetical lifetime earnings (column 2 and 3).[58]

Table 2.6: Cost-Benefit Calculations of Personality Changes on Lifetime Earnings

	Wage return of reform (1)	Change in Lifetime Earnings Absolute (EUR) (2)	Relative (%) (3)
Males	-3.01%	-10081.96	-0.60%
Males from non-intact family	-2.14%	4973.06	0.29%
Males with migration background	-4.08%	-28647.57	-1.70%
Females	-5.74%	-35925.03	-3.08%
Females from non-intact family	-2.28%	5608.51	0.48%
Females with migration background	-4.24%	-17932.84	-1.54%

Source: Own calculations. The first column is obtained through multiplying the reform induced changes in personality traits with the respective following wage returns to a one-standard deviation change in the trait: Males: Openness -0.017, Conscientiousness +0.013, Extraversion +0.012, Neuroticism: -0.088
Females: Openness +0.023, Neuroticism -0.115, Locus of control +0.049
To obtain columns 2 and 3, the resulting wage changes are used to compute lifetime earnings after the reform (hourly wage*week hours*52 weeks*(working years+1)*(1+wage change)) and compare them to lifetime earnings before the reform (hourly wage*week hours*52 weeks*working years). For this, we use average gross hourly wages (Men: EUR 20.20, Women: EUR 15.83) and the average length of working life (Men: 40.2 years, Women: 35.4 years) from Crößmann and Mischke (2016) and assume full-time employment of 40 working hours per week.

It becomes evident from this very simplified analysis that changes in personality traits are rewarded very differently for distinct groups, e.g. males and females, but also across traits, and that this does not need to go along with the social desirability of a trait. The analysis also emphasizes that the reform entails costs as well as benefits, as the wage penalties induced through personality changes following the reform, are not necessarily translated into a total wage decrease over the lifetime. In some cases, the additional year of working life completely offsets the loss and even yields overall increases in lifetime earnings for some groups. It should, however, be noted that this analysis ignores any consequences of the reform on further outcomes and does not consider general equilibrium effects. Its main purpose serves putting effects into perspective in a simplified back-of-the-envelope calculation.

appropriate reward or penalty which differs by gender.

[57]Note that these are aggregate effects: positive effects for example through increases in extraversion and conscientiousness are canceled through negative effects of the increase in neuroticism.

[58]Especially the absolute values depend critically on the choice of parameters for computing these earnings. Note, however, that a change in hourly wage or in the hours worked per week results in the same relative percentage change presented in column 3. Hence, also for women working only part time, the relative change is still valid.

2.6 Sensitivity Analyses

For a causal interpretation of the high school reform effect we have to rule out that any other factors than the ones considered might drive our estimates. This section discusses several sensitivity analyses that we carried out. The corresponding tables are provided in Appendix A.3.

2.6.1 Age Effects

To account for potential age effects in personality, we so far include age and age squared in all estimations. An alternative way is to regress the original score of each personality trait (without standardization) on age and age squared, next to gender (Nyhus and Pons, 2005). The resulting residuals are then age-free and can be used in the second stage as outcome measures. Our results are robust to this procedure (see Table A.10).

2.6.2 Measurement Issues

A major concern when analyzing personality traits is potential measurement error. In particular when including personality traits as *independent* variables, estimated coefficients are biased if the measure of personality traits suffers from a lack of precision. Therefore, existing studies usually correct for potential measurement error (see for example, Heckman et al., 2013; Heineck and Anger, 2010; Zumbühl et al., 2013). However, in this study the *dependent* variable may suffer from measurement error. In this case, estimates are still unbiased as long as the error occurs randomly, but the variance may increase. We argue that in our study it is reasonable to assume that potential measurement error is random and hence uncorrelated with the assignment of students to the treatment or control group. We do not expect students who are affected by the reform to systematically differ in their self-reporting of personality other than differences induced by true reform effects.

One objection may be that after shortening the high school track, students face a closer proximity to the date of graduation when being interviewed at age 17. At the time of the interview students affected by the reform may hence face more temporary stress compared to their counterparts of the control group, which would bias our estimates. However, Bleidorn (2012) shows based on German data that the Big Five personality traits are stable over the two ultimate years of high school, and change only after graduation. Using our sample of adolescents, we investigate whether proximity to the date of graduation may be a source of non-random measurement error by including the time of the interview (measured in quarters of

the year) interacted with the reform effect (see Table A.10).[59] Significance of the reform effect on neuroticism is reduced after the inclusion of the interview quarter dummies and interaction variables. However, we attribute this to the small sample size and to the large number of covariates. Most important, we find only very few differential effects of the interview timing on personality for the control group and the treatment group.[60] Furthermore, the coefficients of the quarter dummies are mostly insignificant and do not follow a clear pattern. Hence, we conclude that there is no measurement error induced by the timing of the interview.

To take into account changes in personality after graduation, we re-estimate the impact of the reform while controlling for individuals having graduated or not (see Table A.10).[61] While the coefficients point to changes in personality after high school graduation, these effects are not precisely estimated. The effects of the high school reform on the personality traits do not differ from the baseline specification.

Another measurement issue is that adolescents in the SOEP are interviewed via a youth questionnaire, which differs from the adult questionnaire. The youth questionnaire has a much larger focus on (current) secondary education and conditions with respect to school, effort and parents' involvement. The presence of these questions in the youth questionnaire should be kept in mind as a potential source of bias when constructing a sample of adolescents *and* adults from the SOEP. To purge our estimates from potential survey-related effects, we restrict our subsample to adolescents. Indeed, the estimation results are comparable to the results from the entire sample, with agreeableness even significantly improved. The increase in neuroticism decreases in size and statistical significance, possibly due to the substantial loss of observations, but remains economically large (see Table A.10).

Finally, the construction of personality measures by simply averaging item scores along each dimension may imply measurement error. Although we do not expect that this is correlated with the assignment of students to the treatment or control group, we conduct a factor analysis to validate this approach. We in fact find that the items which we use to measure personality in our analysis load on specific factors, which correspond to the dimensions of the Five Factor model and to locus of control.[62]

In sum, we conclude that our study does not suffer from bias due to measurement error. Even in the presence of measurement error, the reform effects we find are unbiased and inference

[59] Including the time of the interview in a larger sample of adults (who already graduated from school) shows hardly any effects and no systematic pattern of the interview month. Hence, we can exclude seasonal variation in personality self-ratings.

[60] An exception is the fourth quarter of the year. However, less than 2% of our sample is interviewed in these last three months of the year, such that these coefficients should not be overvalued.

[61] Only 24 percent of the individuals in our sample have already graduated.

[62] Results are available from the authors upon request.

is still valid, as the estimated standard errors provide an upper bound in this case.

2.6.3 Stability of Personality

Another concern is that the personality measures of the students in our sample may be a mere snapshot and not persist over time. To investigate this, we exploit the panel character of the dataset and use a subsample of students for whom a second self-rating of personality is available in the 2013 survey[63] to compare the two measurements: the rank-order correlation coefficients range from 0.42 to 0.60 for the Big Five dimensions.[64] These results are perfectly in line with Specht et al. (2011) for this age group, and only slightly below the rank-order correlations found for adults between 0.64 and 0.75 depending on the trait (Specht et al., 2011). Hence, we have no reason to suspect that the personality of our sample of high school students is exceptionally unstable.

2.6.4 Estimation Model

As mentioned in Section 2.5, the wild cluster bootstrap proposed by Cameron et al. (2008) to account for the small number of clusters yields even lower standard errors (see Table A.11). Hence, this technique completely confirms our findings. Therefore, in all other estimations, we report the more conservative results without bootstrapped standard errors.

As the Likert-type scale for the assessment of personality is an ordinal scale, an ordered probit model may be more appropriate than OLS regressions. This is true if changes of one unit have different implications depending on their location on the scale, that is, if the scale is non-linear. Therefore, we run an ordered probit regression, with the outcome categories being the standardized measures of each personality trait (see Table A.11). This yields very similar results, confirming our earlier findings.

With multiple outcomes, naturally the concern of multiple testing arises. To rule out that our results are found by pure coincidence as a result of running several regressions, we construct a summary index for the Big Five personality inventory. According to Anderson (2008), these summary index tests are robust to overtesting as adding outcomes to the index does not increase the probability of a false rejection. Further, they are potentially more powerful and provide an estimate of a more general effect. In our case, summarizing the Big Five outcomes by weighting each trait to maximize the amount of information captured in the index, reveals a significant

[63] This is the case for 43 percent of our original sample. Naturally, all students observed in 2013 for the first time cannot be considered.

[64] These computations use the non-standardized measures of personality. Rank-order correlation coefficients are measured by Spearman's Rho. Note that information on locus of control is not surveyed in the wave 2013.

reform effect of 0.12 standard deviations on this index (see Table A.11). If locus of control is added as a further trait to this summary index, the increase remains statistically significant at 0.08 standard deviations. This indicates that the reform affected students' personality, in general.

2.6.5 Selectivity

As outlined in Section 2.4.3, we consider selection from the treatment into the control group or vice versa highly unlikely due to relatively high moving costs involved. Indeed, for 92% of the individuals in our sample, we know that they have not changed residence since their childhood.

Still, we consider the possibility that individuals living close to a state border choose to attend school in a different state to avoid the reform. Hence, we define a subset of late-adopter states whose neighboring states have all adopted the reform already.[65] In these states, any self-selection from the treatment into the control group by moving to a neighboring state (or by attending school in the neighboring state if living at the border) to avoid the reform is thus ruled out. We still find a positive and large coefficient on neuroticism (see Table A.12), which is not, however, statistically significant, as can be expected given the sharp reduction of the sample size.

A more severe concern of selectivity is self-selection *out* of the sample, that is, students who would have originally attended high school but because of the reform chose to follow a different secondary school track.[66] This type of selection may occur in two distinct ways: either directly through a different school choice after elementary school, or through a change in high school dropout rates at a later stage. Both is unlikely, as Huebener and Marcus (2015) find that high school entry rates and graduation rates are not affected by the reform. Still, the former in particular concerns attendance of the comprehensive school (*Gesamtschule*) instead, as this allows students to still obtain the same university entrance qualification. However, the restriction of our sample to states where comprehensive schools typically do not exist[67] reveals virtually the same coefficient on neuroticism (see Table A.12). However, the increases

[65] This group contains all states where the first students affected by the reform graduate in 2012 or later: Baden-Wuerttemberg, Bremen, Hesse, North Rhine-Westphalia, Berlin, Schleswig-Holstein, and Brandenburg.

[66] Note, however, that if self-selection out of high school based on personality traits were an issue, we would expect especially those students to leave or not enter high school who are most worried about the increase in the pace of learning and who fear not being able to cope with the higher workload after the reform. This implies that we would underestimate the true negative effects of the reform in terms of stress and emotional instability among high school students, given that the least emotionally stable students have left the treatment group. Our finding of an increase in neuroticism as depicted in Section 2.5 may in this case represent only a lower bound and a conservative estimate of the true impact of the educational reform on personality.

[67] We define states a those where comprehensive schools typically do not exist 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).

in extraversion and agreeableness are even stronger in the states without comprehensive schools than the average effects, which points to a weakening of the reform effects if students find a way to avoid the new system. Lastly, both possible channels of self-selection, either through a change in school choice or a change in high school dropout rates, would not only change the composition of high school students, but also the composition of students enrolled in the other types of secondary schools, which these students would have to attend when opting out of high school. However, findings in Section 2.6.9 on reform effects for students from other school types support the assumption that there are no such changes in the composition of students.

2.6.6 Announcement Effects and Double Cohort

Since in some states the high school reform was a subject of public discussion prior to its implementation and was hence anticipated, one may raise concerns about announcement effects among cohorts of students around the date of implementation. Post-reform students may have tried to skip one grade to switch to the control group. However, this is highly unlikely as skipping a grade is nearly impossible and students would end up graduating in the same year as they would have originally. Likewise, it is very unlikely that pre-reform students repeat a grade on purpose and switch to the treatment group just to catch up with their original cohort at the time of the high school diploma. In turn, it is more plausible that pre-reform high school students who were at risk of repeating a grade may have tried to evade the reform by putting more effort in staying in their original grade, that is, to remain in the control group.

However, these concerns only apply to the last cohort graduating from high school with 13 total years of schooling and the first cohort graduating with 12 years. This double cohort could feature further peculiarities that may lead to confounding effects or may offset true effects of the reform. Therefore, we exclude this double cohort of graduates in an additional model (see Table A.12). This estimation reveals an even stronger increase in neuroticism, and hence supports the notion that reform effects are to some extent offset in the double cohort of graduates.[68]

2.6.7 Implementation Effects

For the evaluation of the new high school regulation, it is of interest whether the effects are persistent even among cohorts entering high school several years after the implementation of the reform. In many states, the reform has been implemented ad-hoc, for example, relevant books and subject matter were not yet adapted to the increased pace of learning. Any effects

[68] This finding provides a possible explanation why Thiel et al. (2014) find smaller effects in their analysis, which is only based on the double cohort of graduates.

on students' personality traits could therefore also stem from the chaotic implementation of the reform.[69] To investigate whether the effects vanish with time that elapsed since the implementation of the reform, we add dummies for the second cohort affected by the reform, and for cohorts three and higher (see Table A.12). The results reveal that for both groups – the second cohort affected, as well as students affected even three or more years after the implementation of the reform – the effects of the reform are mostly not significantly different from those for students in the first cohort affected, which serve as the reference category in this estimation. Still, it seems that extraversion increased by more among the later cohorts, while there is evidence that agreeableness decreased with the time that elapsed since the introduction of the reform. Note that these results should be handled with caution due to the small sample size,[70] which makes it difficult to clearly interpret our results. However, they illustrate that the reform effects we find are not driven solely by the cohorts immediately affected by the reform but instead can be expected to persist. In particular, our finding of an increase in neuroticism proves to be robust.

2.6.8 Other Institutional Changes

Another aspect of the German high school system that has been the subject of substantial recent interest is the existence of standardized exit examinations (*Zentralabitur*). While these examinations have been in place in some federal states since the 1990s or even earlier, most of the remaining states introduced these standardized exams between 2005 and 2008. Since our empirical strategy exploits the variation over time and region, the effects of the high school reform should be isolated from any other policy changes that occur at different times, which is the case for the introduction of central exit examinations. Furthermore, central exit examinations, if newly introduced, affected both earlier cohorts and the later cohorts that were affected by the high school reform. As a consequence, almost the entire sample under consideration has been subjected to central exit examinations. Nonetheless, it may be that the introduction of standardized exit examinations affects students exposed to the high school reform and those not affected by the reform differently. To rule out that this is driving our results, we consider a subsample of states with standardized exit examinations that have been in place for some time.[71] The results (see Table A.13) confirm that the increase in neuroticism as a consequence

[69] Note however, that this would not contradict our findings that personality traits *are malleable* in adolescence and that secondary schooling plays a role in shaping them.

[70] There are only few individuals in each cell, and as such gender balance, for example, is not guaranteed anymore, neither is the East-West distribution similar among the different years elapsed since implementation of the reform.

[71] States with standardized exit examinations in place for some time are Baden-Wuerttemberg, Bavaria, Mecklenburg-West Pomerania, Saarland, Saxony-Anhalt, Saxony, and Thuringia.

of the reform is indeed not driven by the introduction of central exit examinations, as the respective coefficient is even larger in magnitude.

2.6.9 Placebo Estimation

Finally, we run a Placebo regression to rule out that our results are driven by any other factors influencing students' personalities that are unrelated to the high school reform. Hence, we estimate the effect of the high school reform on the personality measures of students of other school types in the exact same manner. As students who attend the lower secondary school (*Hauptschule*) or the intermediate secondary school (*Realschule*) were not exposed to the high school reform, there should not be any measurable effects of the reform on these students' personality traits.[72] Perfectly in line with this expectation, our Placebo estimation shows no reform effects (see Table A.14).

Furthermore, these findings refute the concern of out-of-sample selectivity following the reform: If there was selectivity out of the sample of high school students based on personality after the reform, we would expect to find a change in average personality among students of other school types. As this is clearly not the case, we can rule out that students who were originally aiming to attend high school selected themselves into alternative school types due to the reform.

2.7 Conclusion

There is a growing body of literature on the importance of personality traits as determinants of diverse economic and social outcomes. In studies dealing with the development of personality, there exists a consensus that the interaction of nature *and* nurture determines skill formation and that skills are shaped early in life. Nevertheless, little evidence exists on the impact of important nurturing factors such as education beyond pre-school age.

We provide first evidence on the malleability of personality traits in adolescence through schooling in Germany, and thus add to the scarce and mostly US-focused literature on the impact of education on non-cognitive skills. Using data on adolescents and young adults from the SOEP, we exploit the German high school reform as a quasi-natural experiment and use the variation over time and across states to establish a causal effect of the compression of the curriculum on the Big Five personality measures and on the locus of control.

[72] Of course, these students could be indirectly affected by the reform, e.g., by facing increased competition for apprenticeship positions as their age advantage over the more highly qualified high school graduates was reduced substantially after the reform. However, these effects should not be as strong as direct effects.

Our estimates show that shortening the high school track, which was associated with a compression of the curriculum, caused students on average to be less emotionally stable. This effect was not only statistically significant but also economically meaningful, as neuroticism increased by more than a third of a standard deviation following the reform. This corresponds to an increase of roughly 0.4 points on the seven-point Likert scale and is robust across model specifications and different groups of students, though it is significantly more pronounced among students in East Germany. Suggestive evidence on an increase in agreeableness and a decrease in locus of control was confirmed to be driven especially by students in West Germany and students with better educated parents, and by female students, respectively. Moreover, our estimates point to additional heterogeneous effects of the change in the educational system on other personality traits. The results suggest a significant increase in openness among students from disrupted families and students with migration background, again corresponding to more than 0.4 points on the seven-point Likert scale. Lundberg (2013) finds openness to new experiences to be particularly relevant in determining college graduation among disadvantaged students. Hence, the improvement in this trait for students from non-intact and migrant families may have important implications for reducing inequality at later educational stages. Additionally, students from non-intact families got substantially more extroverted, while students with migration background experienced a large increase in conscientiousness.

Generally speaking, however, it should be noted that the treatment is restricted to students enrolled in academic-track high school. Those in this school track may be assumed to possess more favorable personality traits than others. While this calls the external validity of the precise estimated effects into question, it does not alter our conclusion that personality traits are malleable through schooling in adolescence. The recent introduction of the reform however, has only allowed us to investigate short-term effects so far.[73]

Nevertheless, our findings indicate that, at least in the short-run, the educational system plays a role in shaping adolescents' personality traits. Since the high school reform was designed in a way that did not affect the overall curriculum, potential mechanisms underlying this influence could include the higher annual workload and increased learning intensity of students, the higher accumulated knowledge at the same age, stronger student-teacher or student-student interactions as a result of longer school days, or the change in time allocation away from non-academic activities. We investigated which of these potential mechanisms were at work in shaping adolescents' personalities through schooling by comparing how further outcomes

[73]These personality traits have proven to be rather stable once entering adulthood. However, it would be an interesting topic for future research to see if the impact of the increase in learning intensity indeed leads to persistent differences in personality of students affected and students not affected by the reform in later life.

changed with the reform. Our results point to adverse effects on health, school performance, and emotions following the reform. We find tentative evidence that the higher workload and the stronger student-teacher or student-student interactions due to longer schools days are the driving forces, as there appears to be no effect on leisure-time activities.

We conclude that personality traits remain malleable in adolescence and that their formation is affected by the educational system, most likely through various channels. Hence, despite different institutional frameworks in the US and Germany, which may imply different skill formation processes, the educational setting is important for non-cognitive skills in both countries. The German high school reform was intended to improve students' competitiveness in the international labor market. The deterioration of some non-cognitive skills and the improvement of others may constitute a potentially substantial source of hidden costs and additional benefits. Put differently, our findings may therefore point to the necessity for educational policies to take the impact of educational changes on personality traits into consideration.

Chapter 3

How Does Education Improve Cognitive Skills?

Instructional Time versus Timing of Instruction*

3.1 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 Wößmann, 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., 2006a, 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

*I thank Silke Anger, Stijn Baert, Friederike von Haaren, Susanne Kuger, Henning Lohmann, Bettina Siflinger, C. Katharina Spieß, Stefan C. Wolter, and seminar and conference participants at DIW Berlin, the Annual Conference of the European Society for Population Economics, the Annual Conference of the European Economic Association, the Annual Conference of the Verein für Socialpolitik, the Annual Conference of the Scottish Economic Society, the International Workshop of Applied Economics of Education, the IZA European Summer School in Labor Economics, the Trondheim Workshop on "Education, Skills, and Labor Market Outcomes", the International Young Scholar SOEP Symposium, the Spring Meeting of Young Economists, and the Essen Health Conference for helpful comments and discussions.

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 study, 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.

The next section gives an overview of the theoretical background and the existing literature. Section 3.3 explains the high school reform in more detail and elaborates on potential channels and anticipated effects. Section 3.4 describes the data and Section 3.5 the empirical strategy. Section 3.6 presents the results including empirical evidence on potential channels. In Section 3.7 several sensitivity analyses are conducted to test the robustness of the findings, before Section 3.8 concludes and discusses the implications.

3.2 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:

$$\theta_{t+1} = f_t(\theta_t, I_t, h) \tag{3.1}$$

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.[74] Fluid 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[75]. 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

[74] 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*.

[75] 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.

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 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 (2016) investigate the long-term impact of education in Germany on word fluency among males and females born between 1940 and 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 (2016), 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 (3.1), 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.[76]

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. The two analyses are complementary with each investigating a different mechanism and therefore distinct implications for policy makers. 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 (2016) 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 girls exhibited higher verbal skills and performance, while boys showed higher mathematical and spatial abilities. Although they note that these gender

[76] Existing literature related to any of these particular mechanisms will be elaborated on in Section 3.3.2 when discussing anticipated effects of the reform.

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, 2014b). 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.

3.3 The German High School Reform

3.3.1 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.[77] 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. Starting in 2001, several German states reduced this time at high school by one year, enabling graduation after completing only twelve years of schooling which was desirable for three reasons.[78] First, the earlier graduation aligned with school duration in other countries, therefore

[77] 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.

[78] 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

making German graduates more competitive on the international labor market. Second, the reform was to increase efficiency in the German education system and hence reduce the costs per student. Third, it allowed earlier labor market entry expanding the labor force by one birth cohort. As Western labor markets face major challenges caused by demographic changes with an increasing disparity between the group of active workers and the rising share of an older population, this relieves pressure on the pension schemes. In contrast, opponents of the reform feared that the reduced time in high school would hurt the quality of education. However, the overall curriculum remained unchanged.[79] 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.[80] 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 given in Table 3.1.

3.3.2 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).[81] 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 cognition, especially during adolescence. Even if they hold, further aspects may hinder or offset their positive effects. Therefore, I conduct

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.

[79] From grade five through receiving the *Abitur*, 265 year-week hours must be completed (KMK, 2013). Year-week hours are the number of weekly 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.

[80] 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.

[81] 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.

Table 3.1: Introduction of the Reform by Federal State

Federal State	Implementation of the reform	Graduation of first cohort affected
Saxony[a]	—	—
Thuringia[a]	—	—
Saarland	2001	2009
Hamburg	2002	2010
Saxony-Anhalt	2003	2007
Mecklenburg-West Pomerania	2004	2008
Bavaria	2004	2011
Lower Saxony	2004	2011
Baden-Wuerttemberg	2004	2012
Bremen	2004	2012
Hesse[b]	2004	2012-2014
North Rhine-Westphalia	2005	2013
Berlin	2006	2012
Brandenburg	2006	2012
Schleswig-Holstein	2007	2016
Rhineland-Palatinate[c]	2007	—

[a]Saxony and Thuringia kept the 12-year school system after reunification.
[b]Gradual introduction for students entering high school in 2004/05 (10% of all schools), 2005/06 (60%), and 2006/07 (30%).
[c]In Rhineland-Palatinate, the reform has only been introduced in selected schools so far.
Source: Autorengruppe Bildungsberichterstattung (2010)

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 B.6 in the appendix) and the high school graduates (NEPS Sample, see Table B.9 in the appendix). 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.[82] Few studies similarly investigate the impact of class hours, but on more curriculum-oriented achievement tests: Andrietti (2015) and Huebener et al. (2016) estimate the same reform's effects on test scores of ninth-graders from the Program for International Student Assessment (PISA). Therefore, these students have been affected by the reform for less time compared to my sample of investigation of seventeen year-olds. Still, Andrietti (2015) finds positive effects in reading, mathematical, and science literacy skills, with the first being driven by female students, while Huebener et al. (2016) find that the reform widens the gap in student performance as the high-performing students benefit the most. Using PISA data as well, two other studies find that differences in instructional time explain variation in student performance across countries (Lavy, 2015), and within Switzerland (Cattaneo et al., 2016), respectively. 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

[82] 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.

cost of other subjects. Cortes et al. (2015) analyze a policy change in Chicago Public Schools that doubled the amount of time devoted to 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[83] 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 cognitively 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.[84] An overview of these anticipated effects can be found in Table 3.2.

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 is tested in the first analysis – increases a person's stock of skills at an earlier stage making any investment thereafter even more productive.[85] As a result, students affected by the reform may have acquired higher cog-

[83] 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.

[84] 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 3.6).

[85] As the overall curriculum was not altered, students are expected to have acquired the same knowledge, not necessarily the same level of underlying of skills. Any positive reform effects are therefore consistent with the Cunha and Heckman (2007) skill formation model, even if total investment has not changed, as long as the productivity of investment decreases with age.

Table 3.2: 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

Notes: + (–) denotes an anticipated increase (decrease) in terms of the score of the respective dimension of cognitive skills. 0 denotes no anticipated change.

nitive 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.[86] 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 3.3 summarizes the hypothesized effects.

[86]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.

Table 3.3: 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

Notes: + (–) denotes an anticipated increase (decrease) in terms of the score of the respective dimension of cognitive skills. 0 denotes no anticipated change.

3.4 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.

3.4.1 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.[87] In addition to various individual and household characteristics, including family background and childhood environment, the SOEP includes cognitive potential measures for different subsamples since

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

2006. The cognitive abilities of adolescents aged seventeen[88], 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[89] 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 twelve-year-school system before Germany's reunification, I consider all students in these two states as affected.[90] 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.[91] Lastly, I restrict the sample to those who successfully completed the cognitive assessment test[92] 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. A detailed description of how it is constructed is given in Table B.1 in the appendix.

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.[93] 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:

[88] 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.

[89] 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.

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

[91] 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.

[92] The share of students who specifically refused to take the cognitive test does not significantly differ between the treatment group (3.6%) and the control group (5.8%).

[93] 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).

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 B.5 and B.6 and the development of cognitive skills over time is graphically illustrated in Figure B.1 in the appendix.

Other variables. To account for individual characteristics that may also influence 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[94]. Furthermore I capture a student's previous performance by the teacher's recommendation after elementary school.[95] 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.[96] A detailed description of all variables included in the analysis is depicted in Table B.3 in the appendix.

3.4.2 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

[94]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.

[95]At the end of elementary school, teachers recommend one of the different secondary school tracks to the student's 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.

[96]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.

of Baden-Wuerttemberg, which targeted students at academic-track high school in their final year.[97] 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 *Gymnasium* in 2012/2013 (Malecki et al., 2014) which corresponds almost perfectly to the German average of 34.4%. Per student expenditures at general schools in Baden-Wuerttemberg (6,900 EUR) were very close to the German average (7,100 EUR) in 2013, and even perfectly coincided with the national average of 7,500 EUR when considering high schools only (Schmidt et al., 2016). 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 and students without valid information on background variables.[98] The final sample consists of 2,128 students, of whom 1,113 are affected by the reform. A detailed description of how it is constructed is given in Table B.2 in the appendix.

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.[99] 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

[97] This study 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.

[98] Further, I exclude 2 students who were born in 1990 and 1999 for being strong outliers in the age distribution.

[99] 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 October 2016, are not yet released. Once released, the data will provide an interesting extension to this work that acknowledges the multi-dimensionality of crystallized intelligence.

elements of the matrix, in a total of 3 × 4 cases with 3 × 3 minutes time. For both of these fluid cognitive skill 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 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. A detailed description of all variables included in the analysis is depicted in Table B.4 in the appendix.

3.5 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.

3.5.1 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.[100] However, the amount of education received during

[100]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

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 counterparts at the same age.[101]

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:

$$y_{ist,17} = \alpha \text{REFORM}_{st} + X_i\beta + \sum_s \gamma_s \text{STATE}_s + \sum_t \delta_t \text{YEAR}_t + \varepsilon_{ist}, \tag{3.2}$$

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)[102] 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 with 15 different states.[103]

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

the year they turn seventeen.

[101] 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).

[102] 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.

[103] 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 B.13 in the appendix.

on initial abilities and socio-economic background, I assume 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,[104] 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.[105] Hence, selection *within* the sample, i.e. between treatment and control group, was hardly possible.[106] Still, the robustness checks in Section 3.7 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*.[107] 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.

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,

[104] 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.

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

[106] 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.

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

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.

3.5.2 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.[108] 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,[109] 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 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 (3.2) as there is no variation over time and across

[108] 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.

[109] 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 3.6.1.

states in this sample:

$$y_{ij} = \alpha \text{REFORM}_i + X_i\beta + \varepsilon_{ij}, \tag{3.3}$$

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.[110] 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.[111] However, selection *out of* the sample may have been possible, but is again assumed to be unlikely.[112] Furthermore, Table B.10 in the appendix 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.

[110] 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.

[111] Similar to the discussion on selection in Section 3.5.1, 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.

[112] This could happen if students drop out of this double cohort either by repeating or skipping a class. 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 3.5.1 for a more extensive discussion of potential selection.

3.6 Results

3.6.1 The Impact of Instructional Time on Cognitive Skills

Estimation results of equation (3.2) are presented in Table 3.4.[113] The few salient effects of the individual control variables that reveal statistical significance[114] 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.[115]

It can be seen that, on average, the increased instruction quantity induced by the reform has no significant impact on students' cognitive abilities.[116] Yet, the positive sign across all dimensions is in line with theory. However, when allowing for effect heterogeneity by gender (see Table 3.5), it can be seen that these are driven by improvements among male students: While there is virtually no effect among female students, male students' 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,[117] 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 provides two potential reasons for why instructional time may be more beneficial for male students than for female students. On the one hand,

[113] 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.

[114] 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 B.6 in the appendix.

[115] The reverse could, however, also be plausible to some extent: the sample is special in the sense that it only comprises high school students, and this selection into the highest track may differ between socio-economic groups. This could explain the negative but insignificant coefficients of high parental education in case a much larger share of students in this group make it into high school while it is only the very best among students with low parental education. The effects of the individual characteristics on cognitive skills may, therefore, not be representative for the average adolescent in Germany but are included here merely for control purposes.

[116] 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.

[117] 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 3.4: 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.

the initially higher numerical skills can shape male preferences for choosing mathematically demanding subjects.[118] 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. This is also in line with the finding by Huebener et al. (2016) that the reform widened student

[118]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.

Table 3.5: 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.

achievement gaps as high-performing students benefit the most. 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, 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 socioeconomic 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.[119]

A different type of heterogeneity can arise due to institutional differences. The allocation of the increase in class hours to the different grades is at the discretion of the federal states and even schools. Yet, in the states of Berlin, Brandenburg, and Mecklenburg-West Pomerania there is less scope as high school begins only with grade seven. Excluding these states from the analysis, the positive effect on males' numerical skills is even more pronounced while the

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

gender heterogeneity is slightly reduced (see Table B.11 in the appendix). This suggests that especially the earlier years in high school, i.e. grades five and six, may be critical, and that a larger scope to allocate increases in class hours may be less detrimental for equality. For the evaluation of timing effects in more detail, however, see the results in Section 3.6.2.

Finally, I investigate the persistence of effects in this sample. For a small subsample, an additional measure of cognitive skills is available via the adult questionnaire in 2012. At this point, the individuals are of age 18 to 25, with most of them, thus, having completed high school. The very small number of observations of 80 or less, naturally yields imprecise estimates (see Table B.12 in the appendix). However, the signs of the respective coefficients qualitatively confirm the results with males still benefiting especially compared to females.[120]

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 3.6).

First, I analyze changes in leisure-time activities, as in an experimental study, Schellenberg (2004) shows that e.g. music lessons increase IQ among children. I therefore analyze whether such 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.[121] 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 cope with

[120] In addition to instructional time, these results may be confounded by differential timing and increased instructional time at university or vocational training. For a clean analysis on timing of instruction, therefore, see Section 3.6.2.

[121] 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 3.6: 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.

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.

3.6.2 The Impact of Timing of Instruction on Cognitive Skills

Estimation results of equation (3.3) are presented in Table 3.7.[122] 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.

Table 3.7: 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.

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

[122]Due to a resulting loss in the number of observations, I do not include school characteristics in this and the following estimations. Table B.19 in the appendix shows however that, when further controlling for school characteristics, results are not altered.

Table 3.8: 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.

heterogeneity (see Table 3.8).[123] 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 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

[123] 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.

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 likely is the result of an interaction of this age effect 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.[124]

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.[125]

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 the implementation of the reform went.[126] 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

[124]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.

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

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

Table 3.9: 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.

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 3.9). 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 why the male advantage in crystallized intelligence obtained at the age of seventeen is offset until graduation as male students reduce their cognitively stimulating leisure-time activities compared to females.[127] 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

[127]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.

participation in sports during leisure-time seems to increase as well, at least on average, but the change amounts to only three percentage points.

3.7 Sensitivity Analyses

To confirm the positive effect of the increased instructional time, at least among male students, I conduct several robustness checks. The corresponding tables are provided in Appendix B.4. Table B.13 presents methodological alterations: As mentioned in Section 3.5.1, 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.[128]

To deal with the measurement of cognitive skills (see Table B.14), I first exclude all students aged eighteen or nineteen years from my sample to rule out potential age effects.[129] While the positive impact of instructional time on males' numerical skills is preserved, fluid intelligence scores, however, increase in this specification as well.[130] 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 B.15). 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

[128]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, 2008).

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

[130]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.

are preserved.

Crucial for the interpretation of a causal effect in specification (3.2) is that there is no selectivity. As elaborated on in Section 3.5.1, selectivity is unlikely but possible (see Table B.16). 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: (i) federal states that adopted the reform relatively late and neighbor states where the reform was already implemented[131] and (ii) students who never moved from their place of childhood.[132] Both estimations show that the effects are completely preserved with the 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.[133] 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.[134] 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 B.17: 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

[131] 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.

[132] These constitute 94.5% of my original sample, hence demonstrating very low regional mobility. Place of childhood is defined as the town or area where they lived the majority of their life until age 15.

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

[134] 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).

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[135] 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. 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 (3.2), 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 B.18).

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; the results of which are provided in Appendix B.5. 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 B.20 and B.21): 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

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

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 school fixed effects to the estimation to eliminate student sorting across schools and simultaneously control for all school-specific factors. Again, this reduces statistical significance but the results remain qualitatively similar. Fourth, 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 on 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[136].

3.8 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 study 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

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

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 seven-teen-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, 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.

Chapter 4

Is There a Protective Effect of Education on Mental Health?*

4.1 Introduction

Mental health problems are increasing substantially over time and are prevalent across all countries.[137] According to the World Health Organization (2001), 450 million people worldwide suffered from mental or neurological disorders or from psychosocial problems in 2001. The Global Burden of Disease Study 2010 shows that mental and substance use disorders rank first among all leading causes in explaining years lived with disability and fifth in explaining disability-adjusted life-years (Whiteford et al., 2013). According to Frank and McGuire (2000), each year 4% of the population is affected by the most severe mental disorders including schizophrenia, manic depression, and other forms of major depression. These are highly disabling and include psychoses with severe disturbances in cognitive functioning. In Germany, figures from the BKK Dachverband (2015) show that the incidence of an inability to work due

*This chapter is based on joint work with Daniel Schnitzlein. This project was funded by the Federal Ministry of Education and Research under the title "Nicht-monetäre Erträge von Bildung in den Bereichen Gesundheit, nicht-kognitive Fähigkeiten sowie gesellschaftliche und politiche Partizipation" (FKZ: NIMO-ERT2/Kassenzeichen: 8103036999784/#30857). We thank Falk Voit for excellent research assistance and Aileen Edele, Daniel Kemptner, C. Katharina Spieß, Kristof de Witte, and seminar participants at DIW Berlin, the University of Hamburg, and ZEW Mannheim for helpful comments and discussions.

[137] See Frank and McGuire (2000) for an overview.

to mental and behavioral disorders has more than doubled between 2004 and 2014.[138] In 2002 work absence amounted to 1,088 days per 1,000 insured non-retired individuals. This number more than doubled to 2,395 in 2014, with a substantially larger incidence among women than men.

For the affected individuals and families, mental health problems constitute a severe burden. They not only decrease quality of life by causing severe disabilities, but also increase mortality. Estimates by the World Health Organization (2001) suggest that every year one million individuals commit suicide, with between ten and twenty million attempting it. Generally over 90% of those who commit suicide have a psychiatric diagnosis at the time of death (Bertalote and Fleischmann, 2002). Family members are indirectly affected as they may face economic difficulties, emotional problems and stress in adjusting, disruption of household routines, as well as restrictions on social activities (World Health Organization, 2001). In addition, the increasing prevalence of mental health problems is costly to society both in terms of direct spending on treatment as well as in terms of economic losses due to reduced productivity (Frank and McGuire, 2000). In 2008, mental disorders accounted for the third largest category of medical expense in Germany at EUR 28.7 billion (RKI/Destatis, 2015), thus posing a large burden on the health care system. In addition, social security systems and employers are affected by long periods of work inability or reduced productivity. Bubonya et al. (2016) find that poor mental health is significantly associated with both higher absence rates – absenteeism – and lower productivity while at work – presenteeism. In Germany, the average period of work absence due to mental disorders was 39.1 days in 2014; three times longer than due to other illnesses at 13.3 days (BKK Dachverband, 2015). Furthermore, the 2014 national economic loss was estimated at a loss of production of EUR 8.3 billion and a loss of net product of EUR 13.1 billion (BMAS/BAuA, 2016).

Despite the increasing prevalence of mental health problems and the severe burden on several affected parties, little is known about the causal determinants of mental health conditions. Understanding these is crucial for policy makers in order to adequately address or prevent mental disorders. Education may be one such important determinant that yields beneficial and protective effects on health. In this case, schooling could be used by policy makers to

[138] These numbers are based on data from 78.1% of BKK insured, who represent about 13% of all publicly insured individuals in Germany. In addition to true incidence, trends in diagnose habits over time and a decrease in the stigmatization of mental disorders may account for some of the increased prevalence that is observed.

target mental health and educational interventions would yield important spillover effects. Yet, most of the empirical economic literature is limited only to the impact of education on physical health, establishing important protective effects. However, mental health is largely ignored in this context, not least because it is usually more difficult to measure.

Since the seminal work of Grossman (1972a,b) many contributions present theoretical models for the demand of health.[139] These models treat health capital as one dimension of the human capital stock. While there are different views of whether education raises efficiency of inputs or alters the composition of inputs, the models agree in predicting a positive correlation between both the demand for health and the stock of health capital with the level of education of an individual, even in the absence of a causal effect. All theoretical contributions thereby focus on the physical health aspect. In the subsequent sections, we analyze if these predictions also carry over to the aspect of mental health.

In the empirical economic literature, the relationship between education and *physical* health is widely studied, establishing causal protective effects: Schooling, measured in years or degrees obtained, improves health outcomes, such as long-term illness (Kemptner et al., 2011) and mortality (Lleras-Muney, 2005), as well as health behaviors including absence of smoking (Kenkel et al., 2006). Additionally, education yields positive spillover effects on the next generation's physical health (Currie and Moretti, 2003; Kemptner and Marcus, 2013).

In contrast, however, the causal effect of education on *mental* health is under investigated in the economic literature. The few existing studies can mostly be distinguished into two strands. In the first, studies investigate long-term effects of schooling on old-age cognitive abilities including memory functioning, which relates to dementia. Banks and Mazzonna (2012) find increases in executive functioning among males and in memory functioning among males and females older than 50 when exploiting a compulsory schooling reform in England in a regression discontinuity design. Glymour et al. (2008) study the effect of changes in compulsory schooling laws on US residents using data from the Health and Retirement Study, identifying positive effects on memory, but inconclusive results on cognitive functioning. Based on SHARE and SHARELIFE data, respectively, on several European countries, Schneeweis et al. (2014) and Crespo et al. (2014) use compulsory schooling laws, which vary across countries, as an instrument to estimate positive effects of education on memory functioning of individuals aged

[139] An overview of the theoretical literature and empirical studies testing the hypotheses from the theoretical models can be found in, among others, Grossman (2000, 2006), and Lochner (2011).

50 or older. Additionally, Crespo et al. (2014) find that one extra year of schooling decreases the risk of depression by 6.5 percentage points. This leads to the second strand of the literature, which is concerned with specific mental disorders, in particular depression. In non-clinical surveys, the prevalence of depression is often assessed via a malaise score of eight or higher. The malaise score is computed from a 24-item self-completion inventory developed by Rutter et al. (1970) to assess emotional disturbances and associated somatic symptoms.

Employing propensity score matching among individuals born in 1958 in the United Kingdom, Feinstein (2002) finds that as educational levels increase, there is a reduction in depression likelihood at age 42. Instead, Chevalier and Feinstein (2007) use two variables from the individuals' past to instrument education of respondents of the British National Child Development Survey: smoking at age 16 and the teacher's assessment whether post compulsory schooling would benefit the child. Their IV estimates of 42 year-olds result in a 5 to 7 percentage point decrease in depression for those with a secondary education qualification. However, these effects are non-linear, largest at low to mid-levels of education, and most pronounced for this oldest age group considered.

Other than these disorder-specific studies, only two analyses use a more general measure of mental health. Johansson et al. (2009) use parental education as an instrument among respondents to the Health 2000 population survey in Finland, but find that years of education do not impact their general measure of mental health assessed through the GHQ-12, a 12-item General Health Questionnaire that covers both the inability to perform usual daily tasks as well as the experience of newly arising and distressing symptoms. Investigating tertiary education instead, instrumented through college availability and financial support for university studies, Kamhöfer et al. (2015) also do not find any effects on the Mental Component Summary score of adult respondents of the German National Educational Panel Study (NEPS).

In the present study, we investigate whether the protective effect on physical health generalizes to the mental dimension. In addition to the mere absence of diseases, the World Health Organization (2014) defines mental health as a state of well-being in which individuals can cope with daily life, realize their potential, and contribute to society. To acknowledge this multidimensionality, we analyze the effect of years of schooling on a generic measure of mental health, the Mental Component Summary (MCS) score[140], complemented with self-reported medical in-

[140]Details on this measure are provided in Section 4.4.

formation. To estimate causal effects, we employ an instrumental-variable (IV) technique with two sources of exogenous variation to identify local average treatment effects (LATE) at different parts of the educational distribution: (i) a reform extending compulsory schooling by one year implemented in West German federal states between 1949 and 1969 and (ii) the individual availability of higher education measured by the spatial distance to the nearest university. In addition, we estimate Marginal Treatment Effects (MTE) to infer on effect heterogeneity along the unobserved propensity for taking up education. We use rich individual data on adults aged 50 to 85 from the German Socio-Economic Panel (SOEP) study, augmented by detailed information on universities from the *Stiftung zur Förderung der Hochschulrektorenkonferenz*, the foundation supporting the German Rectors' Conference.

Thus, our study contributes to the existing literature in several ways. First, and most importantly, we apply the MCS as a general measure of an individual's mental health constitution. As it is a continuous scale-measure that is not related to any specific disorder, it allows for identifying even small changes among the general population instead of only focusing on the extensive margin of entering depression. Second, this study is the first to identify causal effects of secondary education on mental health in Germany. As the positive effect on physical health is already established for Germany (see e.g. Kemptner et al., 2011), the results on mental health can be compared and contrasted. Third, our empirical strategy relies on two distinct instruments that allow for estimating LATE at different parts of the educational distribution and for explicitly modeling effect heterogeneity through MTE. This enables us to infer about differential effects and non-linearity with respect to educational levels and individual preferences for education. Lastly, the rich dataset includes self-reported medical information that serves to complement the measure of mental health. Effects both on generic mental health and specific disorders can help understand the discrepancy in the literature: The aforementioned studies on memory functioning and depression find protective effects, while this seems not to hold for mental health in general. It is, however, also of particular interest given that the direction of potential mechanisms is not as clearly positive as for physical health.

Our results support existing evidence on a positive relationship between completed years of secondary schooling and mental health in standard OLS estimations. In contrast, however, the IV estimations reveal no such causal protective effect of mental health. If anything, the estimates point toward a negative effect and we find increasing sleep disturbances, in particular

among individuals with generally lower levels of education. The results are robust to a variety of sample specifications and methodological alterations. Furthermore, they are confirmed when explicitly modeling effect heterogeneity through MTE.

The next section provides a discussion of underlying mechanisms. Section 4.3 explains the empirical strategy and discusses the choice of instrumental variables, while Section 4.4 describes the data. Section 4.5 presents the results, several robustness checks and an analysis of effect heterogeneity. Section 4.6 concludes and discusses the implications of the findings.

4.2 Potential Channels

Education may positively impact health and mental health through several channels. First, the higher income generated through better education may improve healthy living conditions and enable better access to medical care and insurance take-up. Cutler and Lleras-Muney (2010) find this economic argument of *better resources* to account for up to 30% of the relationship between education and health. No need to worry about financial hardships can also directly influence, in particular, psychological aspects of health. Second, education can improve health through *information*: Kenkel (1991), for example, shows that an explicit measure of health knowledge explains at least part of the investment in healthier behavior. Third, *behavior and traits* like self-management may also be direct channels. Currie and Moretti (2003) show that the positive return of education on birth health works, amongst others, through the mothers behaving more responsibly as demonstrated by an increase in prenatal care and a decrease in smoking. In addition, Goldman and Smith (2002) find better treatment adherence among the higher educated with diseases, whereas the lower educated switched treatments more frequently, worsening their health outcomes.

While these channels can explain presumably positive impacts of education on health, the remaining paths are more ambiguous and entail ways through which education can have negative impacts, even on mental health. One is *employment*, as education provides access to higher occupations and positions. While this reduces the risk of unemployment, which is beneficial for mental health (Marcus, 2013), results from the Whitehall II study of British Civil Servants show that respondents in higher occupations have more control over their work and more varied work, although they also report more often having to work at a fast pace (Marmot et al., 1991). While

the larger variation in work tasks may be cognitively stimulating, benefiting cognitive ability, the higher pace may be counterproductive for mental health in general. Indeed, Damaske et al. (2016) find that it is the low socio-economic status workers, as measured by education and income, who report greater happiness at work and lower levels of self-reported and perceived stress.

Another ambiguous channel for education influencing mental health is through *social support*. On the one hand, assortative mating and reduced risk of divorce among the better educated can be expected to be beneficial. On the other hand, Jalovaara (2003) shows that the risk of divorce is increased as females earn more, the likelihood of which rises with education among females. Again, the findings of the Whitehall II study suggest that more individuals with higher occupational status see at least three friends per month than those with lower occupational status; this is reversed when it comes to visiting relatives (Marmot et al., 1991). This might originate from higher mobility rates among the better educated (Malamud and Wozniak, 2012) and as social and emotional support may be crucial when handling psychological challenges and increased work stress, attaining higher occupations could therefore be detrimental.

Lastly, education paired with job scarcity can lead to *overeducation*. Using data on employees from the European Social Survey, Bracke et al. (2013) find diminishing returns to education on mental health with overeducated individuals reporting more depression symptoms.[141]

4.3 Empirical Strategy

To estimate the association between education and mental health, we estimate the following equation:

$$mhealth_{itsb} = \beta_1 + \beta_2 yeduc_{itsb} + \beta_3 X_{it} + \mu_s + \delta_s * b + \nu_b + \eta_t + \varepsilon_{it}, \tag{4.1}$$

where $mhealth_{itsb}$ is the mental health measure of person i, observed in year t, from state s, and born in year b. $yeduc_{itsb}$ are the corresponding years of schooling, X_{it} is a vector of individual characteristics and μ_s, ν_b, η_t are fixed effects accounting for differences between states, years

[141] Here, overeducation is measured via two objective indicators obtained through the realized matches method and the job analyst method.

of birth and years of survey respectively. Additionally, we include linear state specific time trends in the year of birth, $\delta_s * b$. Error terms ε_{it} are assumed to be normally distributed with mean zero and are clustered at the person level i. However, the ordinary least squares (OLS) estimate of β, $\hat{\beta}$, may be biased due to the problem of endogeneity. In our context, endogeneity can arise for several reasons including omitted variables and reverse causality, among others. Important omitted variables that may be correlated with both education and mental health include, for example, parental characteristics and time preferences. Better educated or more involved parents may, at the same time, better support their children's education, but also care more about their health. If individuals themselves have a low discount rate, they value the future more and, therefore, invest both in education and health. In turn, reverse causality occurs whenever mentally healthier individuals obtain a higher education because their better health status eases academic achievement. Both omitted variables and reverse causality lead to specific individuals selecting into higher education based directly on either their mental health condition or factors related to it.

To account for this problem of endogeneity and provide unbiased estimates that can be given a causal interpretation, we employ an IV technique by adding a first stage equation to (4.1) where $yeduc_{itsb}$ is regressed on the instrument z_{itsb} and all exogenous variables from equation (4.1):

$$yeduc_{itsb} = \beta_4 + \beta_5 z_{itsb} + \beta_6 X_{it} + \rho_s + \kappa_s * b + \tau_b + \lambda_t + \omega_{it} \tag{4.2}$$

We insert the fitted value from the regression of equation (4.2) for $yeduc_{itsb}$, $\widehat{yeduc_{itsb}}$, in equation (4.1) and estimate a standard two-stage least squares (2SLS) estimator. The resulting coefficient of interest, $\hat{\beta}_2$, can be interpreted as a local average treatment effect (LATE) for compliers with the instrument. In contrast to the average treatment effect (ATE), LATE only applies to a subgroup of the analyzed sample and, therefore, differs from ATE, especially in the case of non-linear or heterogeneous effects. Thus, the choice of the instrumental variable is crucial for the interpretation of the effects. To identify effects for *different* groups of compliers, we resort to two distinct instrumental variables that are likely not to target the same group of individuals: First, we exploit a reform that expanded compulsory schooling from eight to nine years. This likely only impacts individuals at the lower end of the educational distribution who would have otherwise dropped out of school as early as possible. Second, we measure availability

of higher education by the spatial distance to the nearest university at age 19. A closer proximity may increase study intentions, i.e. especially targeting students at the higher end of education. However, to be an adequate instrument, they must fulfill two requirements: (i) Exogeneity. The instrumental variable must not be affected by the outcome, and it must affect the outcome *only* through its influence on the endogenous variable. Since this assumption is formally untestable, plausible arguments for it to hold are brought forward in the following subsections where our choice set of instrumental variables is discussed in more detail. (ii) Relevance. Variation in the instrumental variable needs to explain sufficient variation in the endogenous variable, i.e. in years of schooling. We follow the recommendation of Staiger and Stock (1997) and provide F-Statistics for exclusion of the instruments in all first stage regressions presented in Section 4.5. In addition to these requirements, the instrumental variable cannot violate the assumption of monotonicity. In our case, this implies that individuals who opt for more years of schooling in the absence of the instrument, would not opt for less if the instrument changes in favor of education. Hence, individuals must either perfectly comply with or not react to the instrument, but need to be assumed not to react *opposite* to it.

4.3.1 Compulsory Schooling Reform as Instrumental Variable

Institutional Background and Change. In Germany, educational policy is the responsibility of each federal state. Common across all states is, however, that children enroll in elementary school at the age of six. Usually after four years, they start secondary school where they chose between three different tracks that are provided at separate schools: *Hauptschule* provides the basic track encompassing eight or nine years after which students leave with a basic school leaving degree. At the *Realschule*, students follow an intermediate track and finish after ten years. After leaving school, students from these two tracks usually start an apprenticeship or vocational training. Only graduation from the *Gymnasium*, the highest track, leads to the *Abitur*, the university entrance qualification.[142] Between 1949 and 1969, compulsory schooling was increased from eight to nine years in the Federal Republic of Germany. As this change in law was implemented by the federal states, the year of implementation differs. Hamburg was the first to raise minimum school leaving years, while Bavaria was the last. An overview over

[142] In some states, comprehensive schools that provide all three tracks at one school exist in addition.

all states is given in Table 4.1.[143]

Table 4.1: Increase of Compulsory Schooling from 8 to 9 Years by Federal State

	First year all students were supposed to graduate after a minimum of 9 years	First birth cohort affected by change in compulsory schooling law
Hamburg	1949	1934
Schleswig-Holstein	1956	1941
Bremen	1958	1943
Lower Saxony	1962	1947
Saarland	1964	1949
North Rhine-Westphalia	1967	1953
Hesse	1967	1953
Rhineland-Palatinate	1967	1953
Baden-Wuerttemberg	1967	1953
Bavaria	1969	1955

Source: Pischke and von Wachter (2005)

Validity of Instrument. This change in law forced students to stay one year longer in school and thus is an ideal instrument for explaining years of schooling. The assumption of exogeneity requires that, conditional on all covariates included in the estimation of equations (4.1) and (4.2), the implementation of the reform is unrelated to any determinants of mental health. These include any unobserved state-characteristics related to the law change that may affect mental health. Therefore, we control for linear state-specific cohort trends. These capture, for example, trends in other policies, including educational expansion, and can also account for macroeconomic factors like economic growth that may increase public resources for education but also improve health through higher living standards. Furthermore, health policy is centrally determined at the national level in Germany, whereas education policy is at the responsibility of each federal state. Hence, residents in all states are subject to the same health system. Any changes therein are, thus, unrelated to the extension of compulsory schooling at different times, and are controlled for by cohort and survey year fixed effects. Figure C.1 in the appendix shows the average years of schooling in the cohorts prior and post reform. The sharp increase of more than 0.5 years with the implementation of the reform illustrates its effectiveness in raising completed years of schooling and reinforces its relevance as instrument. Further, the first-stage F-statistic reported in Table 4.3 exceeds 10, rejecting the hypothesis of weak identification.

[143] A different change in educational policy were the short school years in 1966 and 1967, introduced to move the beginning of the academic year from spring to summer (Pischke, 2007). We address this issue in Section 4.5.2 where we account for this additional change and test the robustness of our findings.

With regard to monotonicity of the instrument, it should be safe to assume that the educational choice of individuals preferring the highest track over both the basic *and* the intermediate track, should be unaffected if the length of the lower one increases but does not exceed the length of the other. However, non-linear preferences of students following the intermediate track may be problematic in case their ideal preference lies between eight and ten years. Before the change, they may have chosen the intermediate track at ten years, but afterwards prefer the basic track with nine years. For this reason, we estimate the impact of the reform on attaining at least an intermediate school leaving degree. The estimate, provided in Table C.7 in the appendix, is statistically insignificant but positive, suggesting that, if at all, attainment of at least an intermediate school leaving degree increases by 4 percentage points. Similarly, we estimate the effect of the reform on dropout probability, to ensure that, although students now spend more years at school, the rate of those completing their education does not fall. The estimated coefficient (see Table C.7 in the appendix) is close to zero and insignificant, suggesting that changes in dropout rates are not a threat to our identification. In addition, this change in compulsory schooling law is used in several studies to estimate causal effects of education on e.g. wages (Pischke and von Wachter, 2008), physical health (Kemptner et al., 2011), and political preferences (Siedler, 2010), where it has been proven a valid instrument.

4.3.2 Minimum Distance to University as Instrumental Variable

Our second instrument for years of schooling is the individual availability of higher education measured by the spatial distance to the nearest university or university of applied sciences.[144] This approach is used, for example, by Card (1995) to estimate earnings returns and by Currie and Moretti (2003) to identify health effect on the newborns of the mothers. In our setting, we compute this spatial distance when the individual is 19 years old as students typically complete secondary education and start tertiary education at this age. Since the place of residence at age 19 is not observed in our data, we resort to the geographic coordinates of the place of birth.[145] Hence, variation in the instrument originates from regional variation and differences across time with new universities opening. Figure C.3 in the appendix illustrates the regional variation in both places of birth and universities, while Figure C.4 in the appendix displays

[144] Compared to universities, universities of applied sciences have a more practical profile and usually grant Bachelor and Master degrees but not doctorates. For ease of notation, we refer to both simply as universities.

[145] We intentionally do not consider any place or change of residence we can observe in the SOEP in year 1984 and later as these are beyond age 19 and, therefore, not exogenous to the educational decisions of the individuals.

the growth of universities over time between 1930 and 1980. To account for non-linearity, we do not use the absolute measure of the distance in kilometers but instead group individuals in ten deciles according to their relative position in the distribution of the distance across the sample.[146] This instrument may induce changes in educational aspirations through different channels, including neighborhood effects and the reduction in transaction costs. Neighborhood effects occur when living close to a university yields a high share of university students in the area, which may stimulate the individual's own educational decision, for example through an information network or through the peer group. Transaction costs are a purely financial argument suggesting that living closer to a university lowers the costs associated with enrolling. For Germany, Spiess and Wrohlich (2010) find that these financial aspects are the main drivers, while neighborhood effects only play a minor role. Importantly, this supports the assumption of monotonicity of this instrument, since rational individuals can be expected to have, not necessarily linear, but monotonic preferences in costs. It is, therefore, unlikely that a closer university offering available education at cheaper costs will induce individuals to attain less education than they would have otherwise.

Validity of Instrument. A threat to the exogeneity of this instrument may be endogenous mobility. To circumvent this problem, we rely on the place of birth of the individual. This is arguably exogenous to any educational choices or health developments of individuals themselves. Yet, it may be endogenous to parental characteristics. In particular, higher educated parents may continue living close to the university they studied at when starting their own family. In general, families who highly value education may sort into areas where local universities exist. Therefore, it is important to control for parental education, which we do in all specifications.[147]

Spiess and Wrohlich (2010) show that in the German context, a shorter distance to the nearest university at the time of completing secondary schooling significantly raises educational aspirations: having a shorter distance than the median, which corresponds to 12.5 kilometers, yields a 7 percentage points higher likelihood of enrolling in tertiary education within five years after school completion. Yet, to enroll in university requires students to complete the highest secondary school track, thus earning the *Abitur*, which is the university entrance qualification.

[146] This accounts for outliers with unusually high distances of up to 148 kilometers. In a robustness check we also use the absolute measure.

[147] Do (2004) discusses this issue in more detail presenting evidence against such sorting. In addition, Spiess and Wrohlich (2010) argue that in the German context moving probabilities are even lower than in the US. Therefore, conditional on parental characteristics, the presence of a university near the place of birth can be assumed to be exogenous.

Therefore, changes in the desire to follow tertiary education need to be accompanied by an increase in secondary school attainment. Therefore, we expect that the spatial distance to the nearest university affects not just tertiary, but also secondary, educational decisions. Indeed, we find a highly significant coefficient for this relationship and, again, the first-stage F-statistic reported in Table 4.3 exceeds 10, rejecting the hypothesis of weak identification.

4.4 Data

The 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.[148] In addition to rich information on individual background, including childhood and family environment, since 2002 the SOEP includes a detailed measure of mental health in every second wave.

Therefore, we only use West German[149] observations from 2002 to 2014 that have valid information on mental health and all control variables. However, as the reform affected cohorts born from between 1933 and 1952, depending on the federal state, onwards, we restrict the sample to observations of individuals aged between 50 and 85 in order to balance the number of observations before and after the reform. We include all available observations per person to maximize the utilization of available information and cluster standard errors at the individual level.[150] We use the resulting sample containing all available observations for descriptive OLS estimations first (OLS Sample). Finally, we construct two separate sub-samples which are best suited for the respective instrumental variable: a Reform Sample and a Minimum Distance Sample. A detailed description of how the samples are constructed is given in Table C.1 in the appendix.

The **Reform Sample** consists only of individuals born within a window of seven years before and seven years after the cutoff birth year for which the longer compulsory schooling

[148] This study uses data from the Socio-Economic Panel (SOEP) for years 1984-2014; version 31.1, SOEP, 2016, doi:10.5684/soep.v31.1.

[149] The restriction to West Germany is necessary when using either instrument. The compulsory schooling reform was only implemented in West Germany, while in East Germany at the time of the German Democratic Republic (GDR), the choice to study was limited and it was impossible to attend a university situated in West Germany. The instrument of the distance to the nearest university is, therefore, less meaningful for individuals growing up in the GDR. Further, the educational systems historically differed between both parts of Germany, such that educational degrees are not directly comparable.

[150] In robustness checks, we address the issue of potential bias due to individuals being observed repeatedly in the panel.

was binding for the first time. To identify these individuals and assign them to the treatment group, which was affected by the reform, or the control group, which was not affected yet, respectively, we further need information on the federal state of schooling. Given that cross-state mobility is generally, and in particular among these older cohorts, low, following Pischke and von Wachter (2008) we use the federal state in which the respondent was observed when first entering the SOEP as a proxy.[151]

In contrast, for the **Minimum Distance Sample** more detailed information on the geographic location is necessary to compute exact distances to existing universities. Ideal would be the place of residence at age 19, when the decision of starting tertiary education is usually made. Any later place of residence may be endogenous if individuals move to and stay where they study. However, since we do not observe individuals at the age of 19, we use the place of birth as a proxy. We use the latitude and longitude coordinates of the center of the municipality where the individual was born.[152] To calculate the distance to the nearest university or university of applied sciences, we use data provided by the *Stiftung zur Förderung der Hochschulrektorenkonferenz*, the foundation supporting the German Rectors' Conference, on the location and opening years of all currently existing universities and universities of applied sciences.[153] For each individual, the resulting measure is the minimum of the distances from his or her place of birth to each university which was existent when the individual was 19 years old. A graphical illustration of the geographical spread of birth places and universities is depicted in Figure C.3 in the appendix. Table C.6 in the appendix shows that both the Reform Sample and the Minimum Distance Sample are comparable in terms of observable characteristics. It is important to note that, by construction, non-German nationality is very low in the Minimum Distance Sample, as only individuals with birth places in Germany are considered, and that the average age differs as the Reform Sample is intentionally restricted in birth years.

Measurement of Mental Health. As the main outcome, we use the Mental Component Summary (MCS) score, which is a continuous-scale measure of general mental health that

[151] In the Minimum Distance Sample, we additionally observe the federal state in which the individual was born. In this sample, we can validate that the for 79% of the individuals, the federal state at survey entry is still the same as at birth, confirming relatively low rates of cross-state mobility. This is in line with findings of Pischke (2007).

[152] The municipality is defined via the municipality key (*Gemeindekennzahl*), with 11,306 different municipalities defined for Germany in 2016. This data was obtained retrospectively in the 2012, 2013, and 2014 waves. For data protection reasons, it is not included in the usual SOEP data distribution, but instead is only available on-site at the SOEP research data center.

[153] All distances are obtained using the geographical information system QGIS and measure the airline distance between the center of the municipality of birth and the visiting address of the university.

can be implemented in large-scale surveys due to its brevity (Ware et al., 1996). It is found to be reliable and valid (Ware et al., 1996) and a psychometrically sound measure of mental health able to detect disorders (e.g. Gill et al., 2007; Salyers et al., 2000). It does not refer to any disorder-specific diagnoses and, therefore, serves as a generic measure of an individual's mental constitution. This measure is used by a few economic studies relating mental health to unemployment (Schmitz, 2011), spousal job loss (Marcus, 2013), and sports (Lechner, 2009). In the SOEP, it is operationalized through the 12-Item Short-Form Health Survey (SF-12), an item battery of 12 questions relating to both physical and mental health (Andersen et al., 2007).[154] For example, questions ask how often in the past four weeks respondents felt down and gloomy or felt that mental health or emotional problems impaired their work or social life. A detailed list of all questions and answer categories is provided in Table C.2 in the appendix. From the answers on these 12 questions, two scores are computed, one for physical (Physical Component Summary, PCS) and one for mental health (MCS). The two scores load on distinct items: PCS loads especially on the dimensions of physical fitness, general health, bodily pain, and role physical, whereas MCS loads primarily on mental health, role emotional, social functioning, and vitality. We standardize the scores to have mean zero and standard deviation one within each sample, with higher values indicating better health, which allows the interpretation of changes in terms of standard deviations and eases comparability to other studies.[155] In addition to these aggregate measures, we complement the analyses with medical diagnoses of depression, dementia, and sleep disorders, which are classified as mental and behavioral disorders according to the International Classification of Diseases (ICD-10)[156].

Other variables. Our primary explanatory variable of interest is education. We measure education by the years of schooling (*yeduc*), which we compute from an individual's obtained secondary schooling degree.[157] Analogous to Pischke and von Wachter (2008) and Kemptner et al. (2011), we assign individuals with no or only a basic degree (*Hauptschulabschluss*) 8 years of schooling if they were not affected by the reform and 9 if they were affected. We count an intermediate (*Realschulabschluss*) or other type of degree with 10 years, while completion of

[154]For details on how the SF-12 version relates to the longer original SF-36 version, see Ware et al. (1996).

[155]Table C.4 in the appendix shows summary statistics of the MCS and PCS scores prior to standardization across the different samples. The original scales are anchored at mean 50 and standard deviation 10 in the 2004 general SOEP population (Nübling et al., 2006).

[156]The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version for 2016 is provided by the World Health Organization on `http://apps.who.int/classifications/icd10/browse/2016/en`.

[157]Thus, individuals who have repeated a grade are not assigned an additional year.

vocational secondary education (*Fachhochschulreife*) is accounted for with 12 years, and only graduation from the highest track (*Abitur*) is considered with 13 years.[158] Further, we control for some individual characteristics including demographic information based on age[159], gender and nationality, socio-economic background measured through parental education, and whether the individual lives in an urban or rural area. We focus on this parsimonious set of controls to avoid including so-called bad controls that may themselves be influenced by education and serve as potential mechanisms. Therefore, with the exception of the area type, which refers to the current place of residence, all variables included are predetermined prior to the individual obtaining education. Additionally, a maximum set of state dummies, year of birth dummies, survey year dummies, and linear birth year state trends, is controlled for in each estimation to account for differences between states and cohorts and for general time trends.

4.5 Results

4.5.1 Main Results

We start the discussion in this section by presenting the results from standard OLS estimations for our aggregate indicators for mental health and, to validate our empirical strategy, for physical health for comparison with the existing literature. In columns 1 and 2 of Table 4.2 results are shown based on all available observations. In total, we observe 45,358 person-year observations from 14,271 persons. On average we observe 3.2 person-years per individual. We pool all the observations and cluster standard errors at the individual level. In column 1 the table presents the MCS score and in column 2 the PCS score. As can be seen, our sample reproduces the well-known positive relationship between education and physical health. The coefficient is positive and highly significant. The OLS estimation also shows a positive and highly significant relationship between education and mental health, which is in line with predictions by the theoretical models. Both estimations control for individual characteristics, birth years, state effects, state trends and survey year effects. These results also extend to the two subsamples we use for our IV analyses: Columns 3 and 4 of Table 4.2 present OLS estimations based on

[158] Years of schooling for individuals who attended comprehensive schools are obtained equivalently according to whether they obtained the basic, the intermediate, or the upper degree at these schools.

[159] Age is implicitly controlled for as dummies for the year of birth and for the survey year are included. In addition, we control for age squared in all estimations to account for potential non-linear age effects.

the Reform Sample and columns 5 and 6 estimations based on the Minimum Distance Sample. The coefficients are perfectly in line across samples with the point estimates between 0.027 and 0.035 standard deviations for MCS and between 0.079 and 0.087 standard deviations for PCS. Yet, these positive correlations might not represent a causal relationship, since important determinants of physical and mental health might be unobserved in the above models.

Table 4.2: OLS Estimations – Effect of Education on Health Indicators

	Outcome variables: aggregate health indicators					
	OLS Sample		Reform Sample		Min. Dist. Sample	
	MCS	PCS	MCS	PCS	MCS	PCS
	(1)	(2)	(3)	(4)	(5)	(6)
yeduc	0.035***	0.079***	0.027***	0.087***	0.032***	0.085***
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
R^2	0.036	0.136	0.032	0.070	0.045	0.128
No. of observations	45358	45358	19185	19185	21426	21426
No. of persons	14271	14271	5789	5789	5250	5250

Notes: SOEPv31.1 waves 2002 to 2014. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. * p<0.1, ** p<0.05, *** p<0.01.

As discussed in Section 4.3, to get closer to the causal effect of education, we make use of two sources of exogenous variation, (i) a compulsory schooling reform and (ii) the individual availability of higher education measured by the spatial distance to the nearest university. Table 4.3 presents the first stage estimations for both approaches.

We present the results from the first approach on the left hand side of the table; on the right are results from the second approach. For each approach, the first stages for MCS and PCS are identical. Our first instrument is the compulsory schooling reform described in Section 4.3.1. As can be seen, the reform increases years of schooling of the individuals by about 0.56 years. The reason that the coefficient is less than 1 is that the sample also includes individuals who would have proceeded with their education irrespective of the reform. Furthermore, the coefficient is in line with first-stage results of existing studies, which find coefficients of the reform on years of schooling ranging between 0.39 and 0.54 using ALLBUS and ForsaBus data (Siedler, 2010) and between 0.58 and 0.7 based on the German Microcensus (Kemptner et al., 2011).

Table 4.3: IV Estimations – Effect of Education on Health Indicators: First Stage Results

	Reform Sample	Minimum Distance Sample
	yeduc	yeduc
	(1)	(2)
Reform	0.558***	
	(0.140)	
Mindist		-0.051***
		(0.011)
F-stat.	16.001	22.374
No. of observations	19185	21426
No. of persons	5789	5250

Notes: SOEPv31.1 waves 2002 to 2014. First stage results estimated by OLS. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

Our second instrument uses the availability of higher education to the individual. As discussed above, we calculate the distance to the nearest university or university of applied sciences when the individual was 19. Using deciles of this measure as instrument, we find a significant negative coefficient in the first stage, which means education levels decrease, the further away (available only at higher costs) a university was when the individual was 19. The coefficient of -0.051 suggests that moving down, for example, five deciles in the distribution of the distance to the nearest university increases schooling by a quarter of a year. This corresponds, for example, to moving from the 6^{th} decile in which students live, on average, around 19 kilometers away from the closest tertiary education institution to the group with the next university at 800 meters.

Table 4.4 shows the results of the second stage estimations. For physical health, we find a significant positive coefficient in the 2SLS estimation using the compulsory schooling reform as an exogenous source of variation (column 2). Thus we find a causal protective effect of education on physical health. This is in line with the findings of Kemptner et al. (2011). Using minimum distance as an instrument for education, results in a virtually similar point estimate; however, it is not significant (column 4).

Turning to mental health, for both instruments we find negative point estimates. Although these are not statistically significant, it is safe to conclude that there is no evidence for a significant protective effect of education on mental health. Using the compulsory schooling

reform as an exogenous source of variation, the point estimate is -0.18 standard deviations (column 1). The 90% confidence interval ranges from -0.399 to 0.034. This upper bound is very small and, hence, even if the effect was positive, it would not be meaningful in size. Using the availability of higher education as our instrument, we find a point estimate of -0.007 standard deviations being very close to zero (column 3). The 90% confidence interval ranges from -0.156 to 0.143.

Table 4.4: IV Estimations – Effect of Education on Health Indicators: Second Stage Results

	Outcome variables: aggregate health indicators			
	Reform Sample		Minimum Distance Sample	
	MCS	PCS	MCS	PCS
	(1)	(2)	(3)	(4)
yeduc	-0.183	0.199*	-0.007	0.136
	(0.132)	(0.121)	(0.091)	(0.093)
R^2	-0.084	0.037	0.040	0.120
No. of observations	19185	19185	21426	21426
No. of persons	5789	5789	5250	5250
F-stat.	16.001	16.001	22.374	22.374
Instrument	Reform	Reform	Mindist	Mindist

Notes: SOEPv31.1 waves 2002 to 2014. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

To summarize, based on the aggregate measures available in the SOEP, we can reproduce the positive protective effect of education on physical health known from the existing literature. In contrast, we find no evidence for a protective effect of education on an aggregate measure of mental health.[160] However, in addition to the aggregate measures, the SOEP offers detailed information on a number of mental disorders. We analyze these in the next section to shed some light on the "black box" MCS score.

Table 4.5 presents results from standard OLS regressions using the probability of ever having been diagnosed by a physician with depression, dementia, or sleep disorders. The last column presents an aggregate indicator that equals one if any of these diagnoses are reported. It is important to note that the MCS score, although an aggregate measure, reacts on a finer scale

[160]In unreported IV regressions we find also no evidence for a protective effect of education on the intra-individual decline in mental health: 2, 4, 6, 8, 10, and 12-year changes in MCS remain unaffected. Results are available from the authors upon request.

to mental problems. Hence, it enables the detection of changes in mental health conditions even at the intensive margin. In contrast, the diagnoses used are binary indicators that switch from zero to one only if an individual exhibits severe symptoms recognized by physicians. Hence, they can be interpreted as the extensive margin.

Table 4.5 shows very low OLS estimates for all cases. As these are linear probability models, there is only a very small correlation between education levels and the probability of being diagnosed with one of the disorders.

Table 4.5: OLS Estimations – Effect of Education on Mental Disorders

	Outcome variables: probability of mental disorders			
	Depression	Dementia	Sleep Dis.	Diagnoses
	(1)	(2)	(3)	(4)
yeduc	-0.002	-0.001***	-0.006***	-0.006***
	(0.001)	(0.000)	(0.002)	(0.002)
R^2	0.016	0.016	0.018	0.035
No. of observations	20311	20311	13876	20311
No. of persons	9685	9685	8233	9685

Notes: SOEPv31.1 waves 2009, 2011, and 2013, all available observations. OLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. * p<0.1, ** p<0.05, *** p<0.01.

Tables 4.6 and 4.7, which show the results from 2SLS estimations, support this finding. Overall we again find no evidence of a causal relationship between education and mental disorders.

One important exception is the coefficient for sleep disorders. Using the compulsory schooling reform, we find a positive coefficient, indicating a negative effect of education on sleep of the individuals. Sleep disorders can have various causes, including physical diseases, but are also frequently encountered along with mental disorders (e.g. John et al., 2005) and may be symptoms of for example enduring stress (Åkerstedt, 2006). Moreover, they compromise an individual's daytime performance (Groeger et al., 2004). Poor sleep quality may also trigger and maintain psychiatric disorders (Brand and Kirov, 2011), thus hindering recovery, and is, therefore, an important component of mental health.

To summarize, when using direct diagnoses as indicators for mental health, we again find no evidence for a protective effect of education on mental health. If any, the evidence would again

Table 4.6: IV Estimations – Effect of Education on Mental Disorders: Reform Sample

	Outcome variables: probability of mental disorders			
	Depression (1)	Dementia (2)	Sleep Dis. (3)	Diagnoses (4)
Second stage results				
yeduc	0.034	-0.001	0.118*	0.093
	(0.046)	(0.006)	(0.068)	(0.059)
R^2	-0.038	0.011	-0.325	-0.168
No. of observations	9605	9605	6525	9605
No. of persons	4317	4317	3729	4317
First stage results on yeduc				
Reform	0.507***	0.507***	0.477**	0.507****
	(0.147)	(0.147)	(0.155)	(0.147)
F-stat.	11.892	11.892	9.503	11.892

Notes: SOEPv31.1 waves 2009, 2011, and 2013, Reform Sample. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

Table 4.7: IV Estimations – Effect of Education on Mental Disorders: Minimum Distance Sample

	Outcome variables: probability of mental disorders			
	Depression (1)	Dementia (2)	Sleep Dis. (3)	Diagnoses (4)
Second stage results				
yeduc	0.021	-0.006	-0.054	-0.016
	(0.028)	(0.004)	(0.038)	(0.035)
R^2	-0.004	-0.008	-0.043	0.034
No. of observations	12050	12050	8938	12050
No. of persons	4914	4914	4883	4914
First stage results on yeduc				
Mindist	-0.045***	-0.045***	-0.042***	-0.045***
	(0.010)	(0.010)	(0.010)	(0.010)
F-stat.	20.339	20.339	18.471	20.339

Notes: SOEPv31.1 waves 2009, 2011, and 2013, Minimum Distance Sample. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

point in the direction of a negative effect among individuals at the lower end of the education distribution.

4.5.2 Robustness Checks

To test the robustness of our findings we conduct several sensitivity analyses. First, we consider general issues that apply to both of our approaches and, hence, use both instruments in separate analyses. Next, we make alterations that are approach-specific and, therefore, deal with each instrumental variable individually.

General Robustness Checks. One major issue in both approaches is measurement error in the years of schooling due to short school years. In 1966 and 1967, most federal states introduced two short school years with length 2/3 of a calendar, to move the beginning of the academic year from spring to summer (Pischke, 2007). Hence, for students enrolled in any grade in these two years, their total length of schooling was usually reduced by 2/3. Given that this change was largely administered at the same time or later, compared to the compulsory school reform, we overestimate the actual years spent at school, especially among students already subject to nine years of compulsory schooling. In addition, in the Minimum Distance Sample, it is problematic if the length of schooling is systematically overestimated only among later-borns, since we exploit variation over time in the opening of universities. Other studies argue that (positive) estimates should, therefore, only provide lower bounds. However, in our case of inconclusive results or even negative coefficients, interpretation is not as simple. Therefore, we adjust the measure of years of schooling to specifically incorporate the actual length of school years. For individuals in Bavaria and Hamburg, which did not introduce a change or introduced one long school year instead, the years of schooling remain unchanged compared to our original measure. For individuals in all other federal states, we subtract 1/3 or 2/3, respectively, depending on whether the individual was affected by one or even two short school years, depending on the year of birth and the choice of secondary school track.[161] The estimations based on this alternative measure for years of schooling are presented in Table 4.8. These indicate that results are unaltered.

Next, we specify two alternative definitions of our samples to account for potentially over-weighting more frequently observed individuals due to the unbalanced structure of our panel:

[161] We base this assignment on the overview provided in Pischke (2007), Table 2.

Table 4.8: IV Estimations – Effect of Education on MCS: Accounting for Short School Years

	Outcome variable: MCS	
	Reform Sample (1)	Minimum Distance Sample (2)
Second stage results		
yeduc[a]	-0.209	-0.007
	(0.153)	(0.095)
R^2	-0.105	0.040
No. of observations	19185	21426
No. of persons	5789	5250
First stage results on yeduc		
Reform	0.489***	
	(0.137)	
Mindist		-0.048***
		(0.010)
F-stat.	12.718	21.324

Notes: SOEPv31.1 waves 2002 to 2014. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.
[a]This measure is corrected for the loss in school duration due to short school years.

(i) we weight each observation by the inverse frequency the individual is encountered in the data; and (ii) we restrict the sample to only one observation per person.[162] The estimation results are presented in Table C.8 in the appendix and confirm that there is no evidence for a causal protective effect of education on mental health.

Compulsory Schooling Reform as IV. Now we turn to sensitivity analyses specific to the reform approach and vary the sample window we consider around the reform. Our main results are based on limiting our sample to 7 cohorts prior and 7 cohorts post the reform in each federal state. Alternatively, we consider narrower and wider windows and provide estimates for samples bounded by 3, 5 and 10 cohorts prior and post reform, respectively, in Table 4.9. While the fewer observations in the smaller sample spans of 3 and 5 years before and after the reform naturally decrease the power of our instrument by reducing the value of the F-statistic, the original negative coefficient of education on MCS is clearly preserved in each of these samples, although none are statistically significant.

Minimum Distance to University as IV. Next, we make some alterations in measuring

[162]We chose the one closest to when the individual is aged 67.5, as it is halfway between ages 50 and 80.

Table 4.9: IV Estimations – Sensitivity Analyses: Reform Sample

	Outcome variable: MCS		
	Window around reform		
	± 3 years	± 5 years	± 10 years
	(1)	(2)	(3)
Second stage results			
yeduc	-0.154	-0.245	-0.160
	(0.248)	(0.207)	(0.111)
R^2	-0.055	-0.178	-0.057
No. of observations	8637	14284	25366
No. of persons	2523	4210	8020
First stage results on yeduc			
Reform	0.458**	0.498***	0.518***
	(0.227)	(0.182)	(0.112)
F-stat.	4.068	7.521	21.399

Notes: SOEPv31.1 waves 2002 to 2014, Reform Sample. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

the spatial distance to the nearest university, which are presented in Table 4.10. First, we compute our minimum distance instrument as the absolute distance in kilometers (km). Originally, we code the distance to the nearest university in deciles with respect to the sample to account for potential non-linearities. The direct measure using simply the kilometers yields a positive but insignificant coefficient (column 1). Yet, the first stage results with a slightly reduced F-statistic confirm that the measure in deciles is the stronger predictor of years of schooling.

Second, we differentiate by type of university. Unlike in our original computation of the spatial distance, we exclude universities (of applied sciences) that do not have the right to award doctorates, as well as arts and music colleges.[163] This strict definition leaves us with only 87 universities, which constitute a share of 29% of all institutions. The resulting coefficient (column 2) is close to zero and is not significant.

Third, we compute the distance to the nearest university using only public universities. These 184 universities make up 61.33% of the 300 universities originally considered. The remaining 116 are universities that are recognized by the state but are either private or religious,

[163] For arts and music colleges, the data does not allow to differentiate them into possessing the right to award doctorates or not. For this reason, we exclude all in this specification.

Table 4.10: IV Estimations – Sensitivity Analyses: Minimum Distance Sample

	Outcome variable: MCS		
	Distance as continuous measure (1)	Only Universities[a] (2)	Only public universities[b] (3)
Second stage results			
yeduc	0.092	0.009	-0.102
	(0.122)	(0.102)	(0.099)
R^2	0.034	0.043	-0.006
No. of observations	21426	21426	21426
No. of persons	5250	5250	5250
First stage results on yeduc			
Mindist (in deciles)		-0.045***	-0.048***
		(0.011)	(0.011)
Mindist (in km)	-0.006***		
	(0.001)		
F-stat.	14.455	17.234	20.179

Notes: SOEPv31.1 waves 2002 to 2014, Minimum Distance Sample. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.
[a]In this sample, only universities in the strict sense, with the right to award doctorates, are included. Universities that do not have the right to award doctorates, universities of applied sciences, as well as arts and music colleges are excluded when computing the distance to the next university at age 19.
[b]In this sample, only public universities are included. Universities which are recognized by the state but under private or clerical sponsorship are excluded when computing the distance to the next university at age 19.

typically charging tuition. Cost saving arguments should, therefore, especially apply in case of nearby *public* universities. Still, there is no indication for such difference, and results remain robust (column 3).

Overall, these sensitivity analyses prove our results robust: In none of the specifications do we see evidence for a protective effect of education on mental health. If at all, we mainly observe negative coefficients, especially in, and throughout all, estimations that use the compulsory schooling reform as an instrument. This may also hint at potential effect heterogeneity along the educational distribution, which we address next.

4.5.3 Heterogeneous Effects

Our estimates so far are restricted to the specific groups of compliers with the two instruments. Therefore, we can infer local effects, LATE; however, we cannot draw conclusions regarding the

other groups or compute an ATE valid for the entire sample. In case of homogeneous effects, these coincide. But this rarely is realistic for what Heckman et al. (2006b) term essential heterogeneity when the individual's selection into the treatment (of education in our case) is related to the expected benefits of the treatment. Differential returns to education may occur for a variety of reasons including differences in the counterfactual outcome or characteristics that make education more beneficial for some than for others, like a higher learning ability or quality of the educational institution. In particular, in case of health, the health condition under absence of the treatment, i.e. a low education, may be very different due to e.g. genetic predisposition. If this is related to the decision on the educational level obtained, gains from education may severely differ across different types of individuals. For this reason, we already employ two distinct instruments to infer on individuals, both in the presumably lowest and the presumably highest part of the educational distribution. The IV estimates suggest that among these, education has no protective effect on mental health. Still, effects could be different for individuals in the middle. Moreover, we cannot clearly define the groups of compliers and, hence, the question on exactly whose treatment effects we estimate remains only vaguely answered. For these two reasons we explicitly model effect heterogeneity using marginal treatment effects (MTE, cf. Heckman and Vytlacil, 2005, 2007). Another advantage of this approach is that we can use both our instruments simultaneously without complicating the interpretation of treatment effects. Therefore, we chose the Minimum Distance Sample for this analysis, as it contains the coordinates of birth places for all individuals and base the assignment to the compulsory schooling reform on the federal state of birth. One caveat of the approach is that it allows only for binary treatments. Therefore, we measure education with a binary indicator of whether an individual obtains above median years of schooling. In the sample, this corresponds to having at least an intermediate degree of 10 years. This also makes the compulsory schooling reform a less powerful instrument as changes from eight to nine years cannot be captured.[164] The results are, therefore, not directly comparable to our original IV estimates, but still should provide important insights into effect heterogeneity.

In the following, we briefly introduce the estimation of MTE. A detailed overview is provided in Cornelissen et al. (2016), whose notation we follow closely. We depart from a Roy model based on the potential outcomes model, in which the outcome under treatment (Y_{1i}) and the

[164] Still, Figure C.2 in the appendix illustrates that also in the Minimum Distance Sample the compulsory schooling reform induced an upwards shift in years of schooling.

outcome without the treatment (Y_{0i}) are modeled as $Y_{1i} = \mu_1(X_i) + U_{1i}$ and $Y_{0i} = \mu_0(X_i) + U_{0i}$, respectively. $\mu_j(X_i)$ is the conditional mean of Y_{ij} given the individual covariates X_i in treatment status j and with deviations U_{ji}. The selection into treatment is modeled as a latent variable discrete choice model:

$$D_i^* = \mu_D(X_i, Z_i) - V_i$$
$$D_i = 1 \text{ if } D_i^* \geq 0, \quad D_i = 0 \text{ otherwise,}$$

where D_i^* is the latent propensity to take the treatment, D_i is the observed choice of taking the treatment or not, and Z_i is one or more variables excluded from the outcome equation but influencing the treatment status (i.e. instruments). V_i is an i.i.d. error term that indicates unobserved heterogeneity in the propensity for the treatment, i.e. it can be interpreted as the unobserved distaste or resistance for the treatment. We can rewrite the condition for selecting the treatment ($D_i^* \geq 0$) and apply the cumulative distribution function of V to obtain probabilities bounded between zero and one:

$$F_V(\mu(X_i, Z_i)) \geq F_V(V_i)$$

We can interpret the left hand side as the propensity score to enter the treatment depending on the observables X_i and the instruments Z_i. The right hand side is then the individual distaste for the treatment that, when exactly equal to the propensity score, makes individuals indifferent. Let us denote the propensity score as $P(X_i, Z_i) \equiv F_V(\mu(X_i, Z_i))$ and the individual distaste as $U_{Di} \equiv F_V(V_i)$. We can now define the MTE at any specific U_{Di} as the treatment effect for those individuals who are just indifferent, i.e. their propensity score $P(X_i, Z_i)$ is equal to U_{Di}:

$$\text{MTE}(X_i = x, U_{Di} = p) = \frac{\partial \text{E}(Y_i | X_i = x, P(Z_i) = p)}{\partial p} \tag{4.3}$$

This estimation is closely related to the standard IV approach and, thusly, requires the same assumptions on the instruments (Cornelissen et al., 2016). IV estimates the relative change in the outcome of interest for all individuals whose treatment status is altered through a shift in the instrument. Hence, it aggregates treatment effects over a certain range of the U_D distribution. In contrast, MTE is defined as the treatment effect at a *particular value* of U_D. It estimates

the infinitesimally small changes in the outcome induced by the instrument and divides them by the instrument-induced changes in the propensity scores.

To estimate equation (4.3), we have to make some functional assumptions to make computation feasible. We model the potential outcomes in a parametric linear way, such that $Y_{0i} = X_i\beta_0 + U_{0i}$ and $Y_{1i} = X_i\beta_1 + U_{1i}$. We assume that the observed and unobserved components in the potential outcomes are additively separable, conditional on $U_D = u_D$. Together with the conditional independence assumption of the instruments (i.e. that the instruments are exogenous conditional on the observables, which is discussed in Sections 4.3.1 and 4.3.2), this restricts the shape of the MTE curve to be independent of the observables. It still allows different intercepts, such that the levels of the curve may differ across values of X, but enables us to present a single MTE curve, evaluated at mean values of X. Equation (4.3) now reduces to:

$$\text{MTE}(X_i = x, U_{Di} = p) = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p}, \tag{4.4}$$

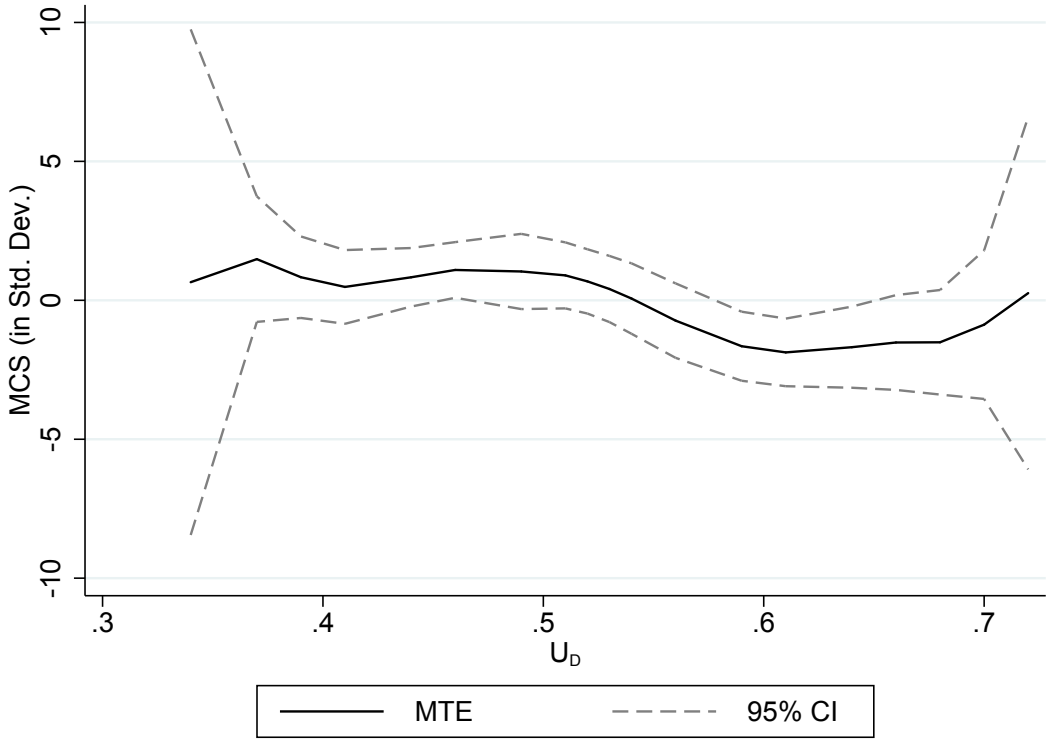
where $K(p)$ is a nonlinear function of the propensity score. To estimate $K(p)$, we chose a semiparametric (Heckman et al., 2006b) polynomial approach with a third order polynomial in the propensity score.[165,166] Importantly, the semiparametric approach allows the estimation of MTE over only the common support of treated and untreated individuals. In our case, this ranges from $U_D = 0.34$ to $U_D = 0.72$, encompassing the entire sample, and is, thus, very informative.

Figure 4.1 presents the estimated MTE curve, evaluated at mean values of X, along with the 95% confidence bands that are obtained through bootstrapping with 500 replications and clustering at the individual level. The curve reveals a downward slope, such that for individuals with lower values of U_D, hence a larger intrinsic motivation for education, we observe positive values. This would suggest selection into gains; however, none of the values are statistically significant. In turn for individuals with U_D greater than 0.55 and, hence, a greater distaste for education, we observe negative values. For some range of individuals around 0.6 of the U_D distribution, these are statistically significant. In addition, we can recover an estimate for the

[165] Alternative estimation approaches are to fit a parametric normal model, a parametric polynomial model, or a fully semiparametric model.

[166] In other MTE studies, usually degrees of two (Carneiro et al., 2011; Cornelissen et al., 2016) are chosen. In our case, different degrees of the polynomial lead to qualitatively similar MTE curves. Therefore, we present results with degree three, as bias is reduced compared to degree two, but variance is still lower than with degree four and higher.

Figure 4.1: Estimated Marginal Treatment Effects



Notes: SOEPv31.1 Minimum Distance Sample. Semiparametric polynomial model with Epanechnikov kernel and of polynomial degree 3. 95% confidence intervals are obtained through bootstrapping with 500 replications and clustering at the individual level.

average treatment effect (ATE) by aggregating the weighted MTE along the U_D distribution. This results in a point estimate of -0.06, which is very small and close to zero.

Overall, we find that the MTE approach confirms our main IV results: In general, education does not seem to have a causal protective effect on mental health. If any, the evidence rather suggests a negative effect, in particular among individuals at the lower end of the educational distribution: the IV estimations based on the compulsory schooling reform show negative coefficients throughout, although insignificant. This is confirmed in the MTE approach where the curve turns negative and partly significant only among individuals with a relatively higher distaste for education.

4.6 Conclusion

Incidence of mental health problems is increasing and globally prevalent. Despite their severe burden on several affected parties, little is known about the causal determinants of mental health conditions. As for physical health, education may yield important protective effects through various channels. Yet, the economic literature largely focuses on the impact of education on physical health, with few studies investigating mental health at the extensive margin by investigating specific disorders.

To analyze changes even at the intensive margin, we provide first evidence on the causal impact of years of secondary schooling on a generic measure of mental health for Germany. Using rich individual data on adults from the SOEP, we exploit two distinct instrumental variables to estimate causal effects at the lower and the upper tail of the education distribution. Our results support existing evidence on a positive relationship between completed years of schooling and mental health in standard OLS estimations. In contrast, however, the IV estimations reveal no such causal protective effect: Unlike physical health, mental health is not significantly altered by education. If at all, point estimates suggest a negative relationship, at least among lower educated individuals. This is analogous when using direct diagnoses on several conditions as indicators for mental health as sleep disturbances increase. Moreover, the pattern is reinforced when explicitly modeling heterogeneous effects: While we find mostly insignificant marginal treatment effects, there is suggestive evidence for negative effects among individuals with a relatively higher distaste for education when they are forced into it.

Our findings are in line with existing literature as we can confirm protective effects on physical health. They also highlight the importance of an adequate measure of mental health that pays tribute to the broad definition. Existing studies find protective effects on single mental conditions, like depression and memory functioning. Related to the latter, our negative, yet insignificant, coefficients on dementia, are in line with these studies. Nevertheless, coefficient estimates on further diagnoses and mental health in general reveal that these are indispensable for understanding the impact of education on mental health, thus helping to complete the picture.

Summarizing, our estimates suggest that education is not an appropriate tool for fostering mental health in the long-run; thus, there is a need for further research on alternative policies.

Understanding the reasons and identifying potential mechanisms in the relationship between education and mental health may unravel potential for intervention. For example, in case work conditions resulting from completing more years of education may be offsetting factors of otherwise potential benefits, specific policies tied to the labor market may be more promising. Additionally, these might have spillover effects, as mental health is impacted also by spousal labor market success (Marcus, 2013). Average zero effects may also originate from important non-linearities in the effects as an additional year of schooling may have differential impacts depending on the overall level. However, we accounted for this by estimating two distinct LATE and applying MTE analysis, which consistently show no evidence for a protective effect.[167] Future research should also consider qualitative aspects of education: Evidence on the benefits of health knowledge on health behaviors (Kenkel, 1991) suggests that returns may be tied to a specific curriculum. While in Germany physical education is an integral part of every student's curriculum, healthy behaviors fostering mental and psychological aspects, such as stress coping mechanisms, are not usually taught at school. This could account for some of the discrepancy in the effects and should be investigated in more detail. Lastly, as females suffer more often from mental health problems than males, gender heterogeneity may be of particular interest in further investigations, as well as replicating the results with younger cohorts that may be exposed to different levels of stress.

[167]Following Winters (2015), we took data on average maternal education in the state-birth-cohort of the individual from the German Microcensus 1976, 1985, 1995, and 2005 to instrument years of schooling along the entire educational distribution. Unfortunately, however, unreliable first-stage results prevented us from adding it as a valid instrument to the analysis.

Chapter 5

Conclusion

5.1 Critical Discussion

This dissertation investigates the impact of different aspects of education on three different dimensions of human capital. Chapter 2 analyzes changes in personality due to an increase in learning intensity among adolescents at high school. As a result, students are less emotionally stable but, at the same time, especially disadvantaged students, are more open, conscious and extroverted. Chapter 3 investigates two mechanisms through which education may influence cognitive skills. The results suggest that, overall, secondary education delivered at high school especially impacts students' crystallized cognitive skills in adolescence through instructional time and not through the age-distinct timing of instruction. However, they also reveal that increasing instructional time aggravates gender differences in numeracy. In contrast to Chapters 2 and 3, Chapter 4 reveals that education, as measured by years of schooling, has, unlike on physical health, *no* causal protective effect on mental health. If any, the estimates point towards a negative effect among the lower educated and individuals with a relatively higher distaste for education.

While the microeconometric techniques applied in the presented chapters are state-of-the-art for identifying causal effects, they hinge on critical assumptions. The analyses presented in Chapters 2 and 3 in particular require the common trend assumption and the absence of selectivity. In turn, the IV identification in Chapter 4 relies on monotonicity, relevance and

exogeneity of the chosen instruments. Therefore, results should be interpreted with some caution, depending on the plausibility of these assumptions. This is extensively discussed in the individual chapters in Sections 2.6, 3.7, and 4.3; with the evidence supporting the validity of the assumptions presented in various sensitivity analyses. Yet, a caveat all chapters share is the relatively small number of observations. In Chapters 2 and 3 the standard errors are, therefore, relatively large, making statistical inference more difficult. Yet, the results provide clear qualitative interpretations. Larger sample sizes, with sufficient number of observations in different cells of the individual characteristics ensured, would also enable better investigation of simultaneous heterogeneities. Alternatively, larger sample sizes would allow to conduct separate analyses, for example, by gender. These are important in case the skill formation processes differ between the two groups and depend differently on the background characteristics. Interaction effects with the reform may thus work differently for male than for female adolescents. Moreover, a larger number of observations would make distributional analyses feasible: quantile regressions can illuminate the location of effects along the outcome distribution. These are particularly informative from an inequality perspective, as they reveal whether only the top percentiles gain whereas the bottom part does not, or whether the lower and upper parts better align and inequality is, thus, reduced. Unfortunately, however, quantile regressions at tail cutoffs like 10% or 90% hardly can identify meaningful effects with the current sample sizes. In contrast, Chapter 4 is based on a substantially larger dataset. However, IV is a very data-consuming technique, such that the IV estimates are more imprecise than the OLS estimates. Further, with more observations MTE could be estimated more sophisticatedly using a fully semi-parametric estimation which allows for even greater flexibility than the specified polynomial semi-parametric approach. Yet, to the best of my knowledge no other German datasets are better suited for my research questions. Although administrative datasets are much larger in size, the intangible outcomes analyzed can clearly be better mapped in surveys. Additionally, survey data sets contain substantially more extensive background information than administrative datasets, which allow to control for and analyze a variety of important individual characteristics. Ideally, it would be desirable to have not only more observations in these survey data sets, but also more observations per person over a wide time span. This would allow to study human capital formation from a true life-course perspective, estimating important dynamics as proposed by e.g. Cunha and Heckman (2007), and identify short as well as long-term effects of education.

For Germany, however, this will only be available with the respective cohorts in SOEP and NEPS growing older while repeatedly and extensively measuring their human capital.

Apart from these methodological caveats, a few issues call for further research. First, all chapters focus on the case of Germany. Whether results are applicable to other countries or different institutional settings remains unanswered. Still, the results provide an important contribution by extending the mostly US-focused literature to a modern developed country with cheap and easy access to education. Further, the quality of education is comparable to that of many other countries, as Germany ranks in the decent middle range in terms of PISA scores but above the OECD average (OECD, 2014a). Hence, the estimation results can serve as a valid benchmark for other countries with similar characteristics. Second, Chapters 2 and 3 only investigate high school students. Although these constitute a large share of students in secondary education in Germany (34.4% of all German secondary students in the 2012/13 academic year, see Malecki et al., 2014), students in this highest track are assumed to possess more favorable skills than others. This calls external validity of the precise estimated effects into question. Still, conclusions on whether and until what age skills remain malleable might be taken as suggestive evidence also for other groups. Furthermore, low-performing students in the analyses conducted may most adequately resemble students not at high school, as they where initially recommended to follow the basic or intermediate track as well and may, hence, at least to some extent, possess similar characteristics. The estimations reveal that these are not affected differently by the reform than other students, which may suggest that results could even be extrapolated to students following different secondary school tracks. Third, in the analysis on instructional timing of Chapter 3, biological age effects cannot be disentangled from reform effects, as by construction students affected by the reform graduate one year younger. This is important to keep in mind when interpreting the results, as it most likely offsets the potential benefits of earlier instructional timing. Unfortunately, therefore, neither of the two can be precisely quantified. Additionally, the chapter relies on two samples that do not originate from the same data source and, consequently, cannot be interpreted as directly linked observations. The ideal dataset would be a panel in which the *same* students are tested in different cognitive domains once at age seventeen while still in school and once at graduation. Unfortunately, this dataset does not exist to the best of my knowledge.[168] Fourth, the analysis in Chapter 4

[168] In the SOEP, measures of cognitive abilities are available for adults as well in form of an ultra-short questionnaire in 2006 and 2012. However, this data is not suitable for replacing the NEPS data on high school graduates

is restricted to specific cohorts and, thus, age range as the compulsory schooling reform only affected cohorts born between 1933 and 1952, depending on the federal state, and later; but suitable data is only available starting in 2002. Therefore, any sample with a balanced number of observations before and after the reform mainly consists of older individuals. Thanks to the panel structure of the data, age and cohort effects can be separately controlled for; yet transfer to younger cohorts is difficult. Furthermore, it would be interesting to investigate effect heterogeneities over the life course in greater detail with a wider age span. Additionally, the instruments used in this chapter allow the interpretation of effects as LATE only at two specific points in the educational distribution. Further heterogeneous effects in comparison to other groups and non-linearities can, therefore, not be detected in the same manner. However, effect heterogeneity is explicitly modeled along the preference for education through MTE estimations showing that no substantial effects on other individuals have been masked.

5.2 Policy Implications and Further Research

Taken together, the different analyses have a variety of policy implications. First, Chapter 2 shows that personality remains malleable in adolescence and that students from different backgrounds react differently to educational changes. For policy makers this implies that although early educational interventions are advertised for yielding the largest benefits, interventions during secondary schooling ages may still be successful in targeting individuals who are otherwise too old. Further, the insights on heterogeneous effects can help to specifically target disadvantaged students: In particular, students from migrant and non-intact families seem to benefit, most likely through longer school days fostering integration, or a more stable environment, respectively. The resulting improvements in openness and conscientiousness may be important contributors to equality of opportunity, as they correlate with tertiary study intentions and success and, therefore, may help to relieve the underrepresentation of students from disadvantaged backgrounds in higher education (Lundberg, 2013; Peter and Storck, 2015).

for two reasons. First, and most importantly, the test was altered between 2006 and 2012 such that only a third of the respondents in 2012 were tested on the same two dimensions as in 2006. For all others, standardization within the intelligence domain and the survey year would be required to make scores comparable across years. Since, however, in 2006 high school graduates were not yet affected by the reform, while in 2012 a substantial part was, the standardization would erase most of the variation originating from the treatment. Second, as the test requires a computer-assisted interview, only a subset of SOEP respondents was tested. Further restricting these to individuals of high school leaving age who took the identical test in either 2006 or 2012 results in too few observations.

Second, Chapter 3 reveals that secondary education delivered at high school improves cognitive skills through the number of class hours and the respective amount of content taught. This provides a new opportunity for policy makers to target cognitive ability, other than just changing overall school duration. However, the findings also point at the necessity of considering differential effects in order to avoid increasing inequality. The increase in instructional time benefits male students only, while female students have difficulties keeping up with the increased learning intensity and require more additional resources. This widens the male-favoring gender gap in numerical skills. In particular in light of efforts to increase female participation in STEM subjects (OECD, 2014b), this seems undesirable and calls for countermeasures supporting female students.

Third, Chapter 4 reveals that formal education is not an appropriate tool to target mental health. Even though simple correlations suggest that education benefits mental health, causal evidence shows that for the presumably highest and, especially, the presumably lowest part of the educational distribution, there is no such protective effect. Other policies may, therefore, be more promising in achieving this aim. Occupational sorting based on education and labor market participation may be offsetting factors in this relationship, as individuals in higher occupations usually face more stress at work (Marmot et al., 1991; Damaske et al., 2016). Therefore, specific policies tied to the labor market, ensuring for example better working conditions or assisting with stress management, may be more appropriate. Additionally, these might have spillover effects, as mental health is impacted also by spousal labor market experiences such as e.g. job loss (Marcus, 2013). Furthermore, future research should abstract from formal education when identifying returns on mental health. Even though formal education seems not to benefit mental health, this does not need to hold for non-formal and informal education. In particular, cognitively and physically stimulating activities as participation in music and sports, but also social networks could be more vital for mental health than solely formal education and should be investigated in more detail.

Considered together, all these policy implications can be seen as recommendations on how to design future interventions to target the specific dimensions of human capital. In this task one should consider the limitations through e.g. the type of students or cohorts analyzed when judging on the transferability of results to other groups. At the same time, the analyses of this dissertation can be used to conduct an evaluation of the already implemented policies.

Since no benefits of the increase in compulsory schooling or the extension of universities on mental health are identified in Chapter 4, I refrain from evaluating these two policies. This does, however, not imply that they do not entail benefits, but only that these may extend to outcomes not considered here.[169] In turn, the impact of the high school reform on skills analyzed in Chapters 2 and 3 can help to evaluate the reform. A thorough cost-benefit analysis is hardly possible to perform for two reasons. First, the lack of reliable data on the increase or reduction in costs due to the policy change, and second, because it requires the assumption that effects are quantifiable. The latter is particularly critical when studying intangible effects as analyzed in this dissertation. One possibility is to quantify results in terms of prospective earnings.[170] Therefore, Section 2.5.3 provides simple back-of-the-envelope calculation which gives an idea of how to put the identified changes in personality into perspective in terms of lifetime earnings. In contrast, Chapter 3 shows that crystallized intelligence remains unaltered by the reform once students graduate and fluid intelligence likely differs only due to biological age differences. Therefore, in an overall cost-benefit calculation of the high school reform taking into account results from both Chapters 2 and 3, changes in cognitive skills need not be considered but the focus remain on personality traits. In this case, the overall cost-benefit calculations reduce to those provided in Table 2.5.3, which emphasize that the reform entails costs as well as benefits, as the wage penalties induced through personality changes following the reform, are not necessarily translated into a total wage decrease over the lifetime. In some cases, the additional year of working life completely offsets the loss and even yields overall increases in lifetime earnings for some groups. It should, however, be noted that this analysis ignores any consequences of the reform on further outcomes and does not consider general equilibrium effects. To perform an adequate cost-benefit analysis, therefore, many further inputs are needed. On the one hand, costs on the input side changed as schools are busier but at the same time host one fewer cohort. On the other hand, there are many other potential outcomes that may be affected by this change. To properly evaluate the reform, therefore, further research is in order.

Of particular interest in this context is the direct effect on wages. However, long-term effects

[169] Most notably, both policies raised the obtained years of education as implied by the first-stage results. Next to this direct benefit, indirect benefits extend e.g. to physical health (Kemptner et al., 2011). However, since this outcome is not the focus of this dissertation, I refrain from further analyses.

[170] However, results may vary or even turn opposite in terms of other dimensions. Prospective earnings, thus, naturally render an incomplete description of overall costs and benefits, and alternatives should be considered in addition.

of the reform cannot be estimated yet. Instead, some studies look at short-term outcomes that may translate into later labor market related decisions or success: Huebener and Marcus (2015) show that, on average, the reform does not result in students completing their high school education one full year earlier. Instead, the average length of time spent in school is only reduced by ten months due to increased grade repetition. Meyer et al. (2015) find that students affected by the reform are more likely to either postpone enrolling in university or not even attend it. These results suggest that the ultimate goal of earlier labor market entry may not be fully achieved. Yet, the reform could pay off if the per student costs of education are reduced due to increased efficiency. Thus, evaluating the reform's impact on various outcomes, including costs at a policy level, as well as further tangible and intangible effects at the individual level, particularly in the long run, is necessary for a thorough evaluation. In addition, the initiatives in some federal states of reverting the reform provide scope for an analysis of opposite effects.[171]

Much more research is also needed for a more thorough understanding of the link between education and mental health. Even though the results of Chapter 4 reveal that, in Germany, years of secondary schooling do not benefit mental health, some further aspects should be considered in future research. Different and additional instrumental variables could extend the analyses to estimate LATE at further points of the educational distribution which differ from the two investigated, and improve the explanatory power of the instruments in the MTE estimation. As the measure of education employed in this chapter is purely quantitative, future research should also consider qualitative aspects of education. Evidence on the benefits of health knowledge on health behaviors (e.g. Kenkel, 1991) suggests that returns may be tied to a specific curriculum. While in Germany physical education is an integrated subject into every student's curriculum, healthy behaviors fostering mental and psychological aspects such as stress coping mechanisms usually are not taught at school. This could account for some of the discrepancy in the effects between physical and mental health and should be investigated in more detail. Moreover, different types of education, including informal and non-formal activities, may be vital in the production of mental health and should be considered in future research. In addition, the effects of education beyond secondary schooling should be investigated. However, the results of Kamhöfer et al. (2015) on college education suggest a similar pattern to the

[171] Lower Saxony changed back to thirteen years of schooling until the *Abitur* with effect in the academic year 2015/16. In other states including Bavaria, Baden-Wuerttemberg, Hesse, North-Rhine Westphalia, and Schleswig-Holstein, schools of both schemes coexist nowadays, such that students and parents have a choice, to a limited extent though.

results presented in this dissertation. Lastly, other datasets with either more detailed medical information or more job-specific characteristics could improve the understanding of relevant mechanisms. In addition, for policy implications, further causal determinants of mental health need to be considered, ranging over a variety of domains including, for example, earlier childhood interventions, conditions in the family and environment, labor market policies, as well as the interaction with physical health.

To conclude, there are many interesting and important research questions on intangible effects in the economics of education literature and on the formation of human capital. With this dissertation I hope to contribute to the discussion on the design of educational policies that foster and facilitate human capital.

Bibliography

Åkerstedt, T. (2006). Psychosocial stress and impaired sleep. *Scandinavian Journal of Work, Environment & Health*, 32(6):493–501.

Almlund, M., Duckworth, A. L., Heckman, J. J., and Kautz, T. (2011). Personality Psychology and Economics. In Hanushek, E., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 4, chapter 1, pages 1–181. Elsevier BV, Amsterdam.

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

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. (2007). Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF12v2. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):171–182.

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, Universidad Carlos III de Madrid.

Anger, S. (2012). Intergenerational transmission of cognitive and noncognitive skills. In Ermisch, J., Jäntti, M., and Smeeding, T. M., editors, *From parents to children. The intergenerational transmission of advantage*, pages 393–421. New York: Russell Sage Foundation.

Anger, S. and Heineck, G. (2010). Do smart parents raise smart children? The intergenerational transmission of cognitive abilities. *Journal of Population Economics*, 23(3):1255–1282.

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. *The Economic Journal*, 122(560):418–448.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(5):9–49.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4):352–365.

Bertalote, J. M. and Fleischmann, A. (2002). Suicide and psychiatric diagnosis: a worldwide perspective. *World Psychiatry*, 1(3):181–185.

Betts, J. R. (2011). The Economics of Tracking in Education. In Hanushek, E., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, chapter 7, pages 341–381. Elsevier BV, Amsterdam.

Bhalotra, S. and Rawlings, S. (2013). Gradients of the Intergenerational Transmission of Health in Developing Countries. *The Review of Economics and Statistics*, 95(2):660–672.

Bhalotra, S. and Rawlings, S. B. (2011). Intergenerational persistence in health in developing countries: The penalty of gender inequality? *Journal of Public Economics*, 95(3-4):286–299.

Bierman, K., Coie, J., Dodge, K., Greenberg, M., Lochman, J., McMahon, R., and Pinderhughes, E. (2010). The effects of a multiyear universal social-emotional program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, 78:156–168.

Björklund, A., Jäntti, M., and Solon, G. (2007). Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents. *The B.E. Journal of Economic Analysis & Policy*, 7(2).

BKK Dachverband (2015). BKK Gesundheitsreport: Langzeiterkrankungen. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, chapter 16, pages 1487–1541. Elsevier.

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.

Bleidorn, W. (2012). Hitting the road to adulthood: short-term personality development during a major life transition. *Personality and Social Psychology Bulletin*, 38(12):1594–1608.

Bloom, D., Gardenhire-Crooks, A., and Mandsager, C. (2009). *Reengaging High School Dropouts: Early Results of the National Guard Youth ChalleNGe Program Evaluation*. New York: MDRC.

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.

BMAS/BAuA (2016). Sicherheit und Gesundheit bei der Arbeit 2014 - Unfallverhütungsbericht Arbeit. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Bracke, P., Pattyn, E., and von dem Knesebeck, O. (2013). Overeducation and depressive symptoms: diminishing mental health returns to education. *Sociology of Health & Illness*, 35(8):1242–1259.

Brand, S. and Kirov, R. (2011). Sleep and its importance in adolescence and in common adolescent somatic and psychiatric conditions. *International Journal of General Medicine*, 4:425–442.

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.

Bubonya, M., Cobb-Clark, D. A., and Wooden, M. (2016). Mental Health and Productivity at Work: Does What You Do Matter? IZA Discussion Papers 9879, Institute for the Study of Labor (IZA).

Büttner, B., Thiel, H., and Thomsen, S. L. (2011). Variation of learning intensity in late adolescence and the impact on noncognitive skills. ZEW Discussion Papers 11-007, Center for European Economic Research (ZEW).

Büttner, B. and Thomsen, S. L. (2015). Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration. *German Economic Review*, 16(1):65–86.

Caliendo, M., Cobb-Clark, D. A., and Uhlendorff, A. (2015). Locus of Control and Job Search Strategies. *Review of Economics and Statistics*, 97(1):88–103.

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.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In Christofides, L., Grant, E., and Swidinsky, R., editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.

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.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781.

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.

Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24(2):365–389.

Cattaneo, M. A., Oggenfuss, C., and Wolter, S. C. (2016). The More, the Better? The Impact of Instructional Time on Student Performance. IZA Discussion Papers 9797, Institute for the Study of Labor (IZA).

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

Chevalier, A. and Feinstein, L. (2007). Sheepskin or Prozac: The Causal Effect of Education on Mental Health. UCD Geary Institute Discussion Paper Series 200715, Geary Institute, University College Dublin.

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.

Coneus, K., Gernandt, J., and Saam, M. (2011). Noncognitive Skills, School Achievements and Educational Dropout. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 131(4):547–568.

Coneus, K. and Spiess, C. K. (2012). The intergenerational transmission of health in early childhood - Evidence from the German Socio-Economic Panel Study. *Economics and Human Biology*, 10(1):89–97.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.

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.

Costa, P. T. and McCrae, R. R. (1994). Stability and change in personality from adolescence through adulthood. In Halverson, C., Kohnstamm, G., and Martin, R., editors, *The developing structure of temperament and personality from infancy to adulthood*, pages 139–150. Hillsdale, NJ: Erlbaum.

Crespo, L., López-Noval, B., and Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, 43(C):36–46.

Crößmann, A. and Mischke, J. (2016). *Arbeitsmarkt auf einen Blick*. Statistisches Bundesamt, Wiesbaden.

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.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

Currie, J. and Moretti, E. (2003). Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. *The Quarterly Journal of Economics*, 118(4):1495–1532.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1–28.

Dahmann, S. (2015). How Does Education Improve Cognitive Skills? Instructional Time versus Timing of Instruction. SOEPpapers on Multidisciplinary Panel Data Research 769, DIW Berlin, The German Socio-Economic Panel (SOEP).

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).

Damaske, S., Zawadzki, M. J., and Smyth, J. M. (2016). Stress at work: Differential experiences of high versus low SES workers. *Social Science & Medicine*, 156(C):125–133.

Dee, T. and West, M. (2008). The Non-Cognitive Returns to Class Size. NBER Working Papers 13994, National Bureau of Economic Research, Inc.

Do, C. (2004). The effects of local colleges on the quality of college attended. *Economics of Education Review*, 23(3):249–257.

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

Duckworth, A. L. and Seligman, M. E. P. (2005). Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents. *Psychological Science*, 16(12):939–944.

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.

Feinstein, L. (2002). Quantitative Estimates of the Social Benefits of Learning, 2: Health (Depression and Obesity). *Wider Benefits of Learning Research Report No. 6*, London: Centre for Research on the Wider Benefits of Learning.

Fletcher, J. and Schurer, S. (2015). Childhood Origins of Adulthood Noncognitive Skills: The Role of Chronic Health Problems and Exposure to Maltreatment. Life course center working paper series, Life Course Center.

Frank, R. G. and McGuire, T. G. (2000). Economics and Mental Health. In Culyer, A. and Newhouse, J., editors, *Handbook of Health Economics*, volume 1, Part B, chapter 16, pages 893–954. Elsevier BV, Amsterdam.

Gerlitz, J. and Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Research Notes 4, DIW Berlin.

Gill, S. C., Butterworth, P., Rodgers, B., and Mackinnon, A. (2007). Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population. *Psychiatry Research*, 152(1):63–71.

Glymour, M., Kawachi, I., Jencks, C., and Berkman, L. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*, 62(6):532–537.

Goldman, D. and Smith, J. P. (2002). Can patient self-management help explain the SES health gradient? *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10929–10934.

Groeger, J. A., Zijlstra, F. R. H., and Dijk, D.-J. (2004). Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. *Journal of Sleep Research*, 13(4):359–371.

Grossman, M. (1972a). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255.

Grossman, M. (1972b). *The Demand for Health: A Theoretical and Empirical Investigation*. Columbia University Press, New York.

Grossman, M. (2000). The Human Capital Model. In Culyer, A. and Newhouse, J., editors, *Handbook of Health Economics*, volume 1, Part A, chapter 7, pages 347–408. Elsevier BV, Amsterdam.

Grossman, M. (2006). Education and Nonmarket Outcomes. In Hanushek, E. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1, chapter 10, pages 577–633. Elsevier BV, Amsterdam.

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

Heckman, J. J., Pinto, R., and Savelyev, P. (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6):2052–86.

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

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006b). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Heckman, J. J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6, Part B, chapter 71, pages 4875–5143. Elsevier BV, Amsterdam.

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., Kuger, S., and Marcus, J. (2016). Increased Instruction Hours and the Widening Gap in Student Performance. Discussion Papers of DIW Berlin 1561, DIW Berlin, German Institute for Economic Research.

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 Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development*, 73(2):509–527.

Jalovaara, M. (2003). The joint effects of marriage partners' socioeconomic positions on the risk of divorce. *Demography*, 40(1):67–81.

Jamison, E. A., Jamison, D. T., and Hanushek, E. A. (2007). The effects of education quality on income growth and mortality decline. *Economics of Education Review*, 26(6):771–788.

Johansson, E., Böckerman, P., Martelin, T., Pirkola, S., and Poikolainen, K. (2009). Does Education Shield Against Common Mental Disorders? Discussion Papers 1202, The Research Institute of the Finnish Economy.

John, K. and Thomsen, S. L. (2014). Heterogeneous Returns to Personality - The Role of Occupational Choice. *Empirical Economics*, 47(2):553–592.

John, U., Meyer, C., Rumpf, H.-J., and Hapke, U. (2005). Relationships of psychiatric disorders with sleep duration in an adult general population sample. *Journal of Psychiatric Research*, 39(6):577–583.

Kamhöfer, D. A. and Schmitz, H. (2016). Reanalyzing Zero Returns to Education in Germany. *Journal of Applied Econometrics*, 31(5):912–919.

Kamhöfer, D. A., Schmitz, H., and Westphal, M. (2015). Heterogeneity in marginal non-monetary returns to higher education. Ruhr Economic Papers 591, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

Kemptner, D., Jürges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, 30(2):340–354.

Kemptner, D. and Marcus, J. (2013). Spillover effects of maternal education on child's health and health behavior. *Review of Economics of the Household*, 11(1):29–52.

Kenkel, D., Lillard, D., and Mathios, A. (2006). The Roles of High School Completion and GED Receipt in Smoking and Obesity. *Journal of Labor Economics*, 24(3):635–660.

Kenkel, D. S. (1991). Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, 99(2):287–305.

King, Jr., M. L. (1947). The Purpose of Education. In Carson, C., Luker, R., and Russell, P. A., editors, *The Papers of Martin Luther King, Jr. Volume I: Called to Serve, January 1929-June 1951*. University of California Press at Berkeley and Los Angeles.

KMK (2013). Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II. Beschluss der Kultusministerkonferenz (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany). Original July 7, 1972, revised February 7, 2013.

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.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *The Economic Journal*, 125(588):F397–F424.

Lechner, M. (2009). Long-run labour market and health effects of individual sports activities. *Journal of Health Economics*, 28(4):839–854.

Lehrl, S., Merz, J., Burkhard, G., and Fischer, S. (1991). *Mehrfachwahl-Wortschatz-Intelligenztest, MWT-A*. Göttingen: Hogrefe.

Li, J. and Powdthavee, N. (2014). Does Increasing Schooling Improve Later Health Habits? Evidence from the School Reforms in Australia. Melbourne Institute Working Paper Series wp2014n01, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *Review of Economic Studies*, 72(1):189–221.

Lochner, L. (2011). Nonproduction Benefits of Education: Crime, Health, and Good Citizenship. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 4, chapter 2, pages 183–282. Elsevier BV, Amsterdam.

Lundberg, S. (2013). The College Type: Personality and Educational Inequality. *Journal of Labor Economics*, 31(3):421–441.

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.

Malamud, O. and Wozniak, A. (2012). The Impact of College on Migration: Evidence from the Vietnam Generation. *Journal of Human Resources*, 47(4):913–950.

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.

Marcus, J. (2013). The effect of unemployment on the mental health of spouses - Evidence from plant closures in Germany. *Journal of Health Economics*, 32(3):546–558.

Marmot, M., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., Marmot, M., and Smith, G. (1991). Health inequalities among British civil servants: the Whitehall II study. *The Lancet*, 337(8754):1387–1393.

Martins, P. S. (2010). Can Targeted, Non-Cognitive Skills Programs Improve Achievement? Evidence from EPIS. IZA Discussion Papers 5266, Institute for the Study of Labor (IZA).

McCrae, R. R. and Costa, P. T. (1999). A Five-Factor Theory of Personality. In Pervin, L. A. and John, O. P., editors, *Handbook of personality: Theory and research*, chapter 5, pages 139–153. Guilford Press, New York, 2nd edition.

Meyer, T., Thomsen, S. L., and Schneider, H. (2015). New Evidence on the Effects of the Shortened School Duration in the German States: An Evaluation of Post-Secondary Education Decisions. IZA Discussion Papers 9507, Institute for the Study of Labor (IZA).

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research, New York.

Mulligan, C. B. and Sala-i Martin, X. (2000). Measuring Aggregate Human Capital. *Journal of Economic Growth*, 5(3):215–52.

Nübling, M., Andersen, H. H., and Mühlbacher, A. (2006). Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF 12 (Algorithmus). Data Documentation 16, DIW Berlin, German Institute for Economic Research.

Nyhus, E. K. and Pons, E. (2005). The effects of personality on earnings. *Journal of Economic Psychology*, 26(3):363–384.

OECD (2014a). PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science. Organisation for Economic Co-operation and Development (OECD), Paris.

OECD (2014b). 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.

Peter, F. H. and Storck, J. (2015). Personality Traits Affect Young People's Intention to Study. *DIW Economic Bulletin*, 5(1+2):3–9.

Pfeiffer, F., Anger, C., Fitzenberger, B., Gebel, M., Kleinert, C., Maaz, K., Schneider, K., Riphahn, R., and Wößmann, L. (2015). *Bildungsindikatoren im internationalen Vergleich: Bestandsaufnahme und kritische Würdigung*. ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Piatek, R. and Pinger, P. (2016). Maintaining (Locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models. *Journal of Applied Econometrics*, 31(4):734–755.

Pischke, J.-S. (2007). The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years. *Economic Journal*, 117(523):1216–1242.

Pischke, J.-S. and von Wachter, T. (2005). Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation. IZA Discussion Papers 1645, Institute for the Study of Labor (IZA).

Pischke, J.-S. and von Wachter, T. (2008). Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation. *The Review of Economics and Statistics*, 90(3):592–598.

Plug, E. and Vijverberg, W. (2003). Schooling, Family Background, and Adoption: Is It Nature or Is It Nurture? *Journal of Political Economy*, 111(3):611–641.

Proctor, S. P., White, R. F., Robins, T. G., Echeverria, D., and Rocskay, A. Z. (1996). Effect of

overtime work on cognitive function in automotive workers. *Scandinavian Journal of Work, Environment & Health*, 22 (2):124–132.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, 22(9):1325–1343.

Quis, J. S. (2015). Does higher learning intensity affect student well-being? Evidence from the National Educational Panel Study. BERG Working Paper Series 94, Bamberg University, Bamberg Economic Research Group.

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

RKI/Destatis (2015). Gesundheit in Deutschland. Berlin: Gesundheitsberichterstattung des Bundes gemeinsam getragen von RKI und DESTATIS.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., and Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science*, 2(4):313–345.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs*, 80(1):1–28.

Rutter, M., Tizard, J., and Whitmore, K. (1970). *Education, health and behaviour*. Longman.

Salm, M. and Schunk, D. (2012). The Relationship Between Child Health, Developmental Gaps, And Parental Education: Evidence From Administrative Data. *Journal of the European Economic Association*, 10(6):1425–1449.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J., and Osher, F. C. (2000). Reliability and Validity of the SF-12 Health Survey among People with Severe Mental Illness. *Medical Care*, 38(11):1141–1150.

Schellenberg, E. G. (2004). Music Lessons Enhance IQ. *Psychological Science*, 15(8):511–514.

Schmidt, A., Eichstädt, H., and Fußmann, M. (2016). *Bildungsausgaben – Ausgaben je Schülerin und Schüler 2013*. Statistisches Bundesamt, Wiesbaden, Bielefeld.

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics*, 18(1):71–78.

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

Schneider, S. L. (2010). Nominal comparability is not enough: (In-)equivalence of construct validity of cross-national measures of educational attainment in the European Social Survey. *Research in Social Stratification and Mobility*, 28(3):343 – 357.

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.

Schurer, S., Kassenboehmer, S. C., and Leung, F. (2015). Do Universities Shape Their Students' Personality? IZA Discussion Papers 8873, Institute for the Study of Labor (IZA).

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., and Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. Ypsilanti: High/Scope Press.

Siedler, T. (2010). Schooling and Citizenship in a Young Democracy: Evidence from Postwar Germany. *Scandinavian Journal of Economics*, 112(2):315–338.

Specht, J., Egloff, B., and Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101(4):862–882.

Specht, J., Egloff, B., and Schmukle, S. C. (2013). Everything Under Control? The Effects of Age, Gender, and Education on Trajectories of Perceived Control in a Nationally Representative German Sample. *Developmental Psychology*, 49:353–364.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Spiess, C. K. and Wrohlich, K. (2010). Does distance determine who attends a university in Germany? *Economics of Education Review*, 29(3):470–479.

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.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.

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.

Thiel, H., Thomsen, S. L., and Büttner, B. (2014). Variation of learning intensity in late adolescence and the effect on personality traits. *Journal of the Royal Statistical Society Series A*, 177(4):861–892.

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.

Ware, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care*, 34(3):220–233.

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

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J., and Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet*, 382(9904):1575–1586.

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, volume 21 of *Advances in Applied Developmental Psychology*, chapter 4, pages 93–124. Ablex Publishing, Westport, CT, US.

Winters, J. V. (2015). Estimating the returns to schooling using cohort-level maternal education as an instrument. *Economics Letters*, 126(C):25–27.

World Health Organization (2001). The world health report 2001 - Mental Health: New Understanding, New Hope. Report, World Health Organization, Geneva.

World Health Organization (2014). Mental health: a state of well-being. Available on http://www.who.int/features/factfiles/mental_health/en/, World Health Organization.

Wößmann, L. (2003). Specifying Human Capital. *Journal of Economic Surveys*, 17(3):239–270.

Zumbühl, M., Dohmen, T., and Pfann, G. A. (2013). Parental Investment and the Intergenerational Transmission of Economic Preferences and Attitudes. IZA Discussion Papers 7476, Institute for the Study of Labor (IZA).

Appendix A

Appendix to Chapter 2

A.1 Sample and Variables

Table A.1: Sample Construction

Sample Description	Remaining observations
All adolescent respondents in waves 2006 to 2013 and adult respondents up to age 21 in waves 2005, 2009, and 2013 with valid information on federal state, year and month of birth	4,411
Drop students from school tracks other than high school or without information on track choice	1,590
Drop students from Rhineland-Palatinate	1,521
Drop students from Hesse entering high school in 2004 and 2005	1,490
Drop students who repeated one or more grades	1,422
Drop students without information on all Big Five personality traits	1,342
Drop students without valid information on background variables	1,316
Final Sample for Big Five personality traits	1,316
Final Sample for Locus of Control (subsample of the above[a])	1,202

Notes: SOEPv30 waves 2005 to 2013, youth and adult questionnaires.
[a]For adolescents, locus of control was observed in the same waves as the Big Five. For adult respondents it was observed only in waves 2005 and 2010. Therefore, for adult respondents from wave 2009 locus of control was merged from wave 2010, if available. For adult respondents from 2013 locus of control is not available. For these reasons, the sample to investigate locus of control is a subsample of the sample to investigate the Big Five.

Table A.2: Big Five Personality Traits

I see myself as someone who...	
is original, comes up with new ideas	Openness to experience
values artistic experiences	Openness to experience
has an active imagination	Openness to experience
does a thorough job	Conscientiousness
does things effectively and efficiently	Conscientiousness
tends to be lazy (reversed)	Conscientiousness
is communicative, talkative	Extraversion
is outgoing, sociable	Extraversion
is reserved (reversed)	Extraversion
is sometimes somewhat rude to others (reversed)	Agreeableness
has a forgiving nature	Agreeableness
is considerate and kind to others	Agreeableness
worries a lot	Neuroticism
gets nervous easily	Neuroticism
is relaxed, handles stress well (reversed)	Neuroticism

Notes: SOEPv30 waves 2005 to 2013, youth and adult questionnaires.

Table A.3: Internal Locus of Control

How my life goes depends on me.
If I run up against difficulties in life, I often doubt my own abilities (reversed).
Compared to other people, I have not achieved what I deserve (reversed).
What a person achieves in life is above all a question of fate or luck (reversed).
I frequently have the experience that other people have a controlling influence over my life (reversed).
The opportunities that I have in life are determined by the social conditions (reversed).
I have little control over the things that happen in my life (reversed).

Notes: SOEPv30 waves 2005 to 2013, youth and adult questionnaires.

Table A.4: Description of Variables

Variable	Description
Openness	Standardized Big Five measure for openness to experience
Conscientiousness	Standardized Big Five measure for conscientiousness
Extraversion	Standardized Big Five measure for extraversion
Agreeableness	Standardized Big Five measure for agreeableness
Neuroticism	Standardized Big Five measure for neuroticism
Locus of control	Standardized measure for internal locus of control
Age	Age (in years)
Female	Dummy for female
East	Dummy for East German states
Rural area	Dummy for having lived most of the childhood until age 15 in rural area
Non-intact family	Dummy for not having lived with *both* parents for the entire time up to age 15
Working-class father	Dummy for father having blue-collar occupation when student is aged 15, reference category encompasses *all* others
High parental education	Dummy for at least one of an individual's parents having an upper secondary school degree or higher
Working mother	Dummy for working mother (both full-time and part-time) when student is aged 10
Migration background	Dummy for student with migration background
Low-performing student	Dummy for having received a recommendation for a different type of secondary school (*other* than high school) after the fourth grade
High school diploma	Dummy for having graduated from high school
Health	Dummy for health status being good or very good (as opposed to satisfactory, not so good, bad)
Sadness	Dummy for frequency of being sad in the past 4 weeks being sometimes, often, or very often (as opposed to seldom or very seldom)
Satisfaction School	Dummy for satisfaction with overall school performance of at least 9 (on scale from 0, low, to 10, high)
Satisfaction Language	Dummy for satisfaction with school performance in the subject German literature of at least 9 (on scale from 0, low, to 10, high)
Music	Dummy for being musically active
Sport	Dummy for doing sports

Notes: SOEPv30 waves 2005 to 2013.

A.2 Summary Statistics

Table A.5: Summary Statistics of (non-standardized) Personality Traits

	Observations	Mean	Std.Dev.	Mean		Equality of Means
				Control	Treatment	t-stat
Openness	1316	4.847	1.090	4.877	4.788	1.401
Conscientiousness	1316	4.962	1.100	4.990	4.907	1.303
Extraversion	1316	5.010	1.202	5.042	4.949	1.323
Agreeableness	1316	5.403	0.891	5.386	5.436	-0.969
Neuroticism	1316	3.788	1.168	3.718	3.925	-3.045
Locus of Control	1202	5.002	0.798	5.004	4.998	0.124

Notes: SOEPv30 waves 2005 to 2013, sample: adolescents and young adults aged 17 to 21 attending high school or with a high school diploma.

Table A.6: Summary Statistics of Individual Characteristics

	Mean		Equality of Means
	Control	Treatment	t-stat
Age	18.134	17.466	8.264
Female	0.544	0.520	0.809
East	0.118	0.435	-13.919
Rural area	0.268	0.316	-1.841
Non-intact family	0.192	0.209	-0.714
Working-class father	0.202	0.229	-1.110
High parental education	0.577	0.531	1.579
Working mother	0.682	0.771	-3.414
Migration background	0.174	0.148	1.183
Low-performing student	0.122	0.072	2.814
Observations	870	446	

Notes: SOEPv30 waves 2005 to 2013, sample: adolescents and young adults aged 17 to 21 attending high school or with a high school diploma.

A.3 Estimation Results

Table A.7: Timing of Reform Implementation and State Characteristics

	Late Implementation[a] (OLS)	Year of Implementation[b] (Ordered Probit)
Proportion of high school students	-0.035	-0.096
	(0.030)	(0.062)
Conservative government	0.100	-0.939
	(0.335)	(0.665)
Next Election in 2001 or 2002	-0.116	-0.309
	(0.327)	(0.640)
Median population age	-0.084	-1.042**
	(0.170)	(0.411)
GDP per capita	0.000	-0.000
	(0.000)	(0.000)
Constant	4.476	
	(7.037)	
R^2	0.219	
Pseudo R^2		0.230
Observations	15	15

Notes: The regressors are pre-reform state characteristics in the year 2000 from administrative data sources. Proportion of high school students (in percentage points)/median population age (in years)/GDP per capita (in euros): Federal Statistical Office; governing party/election dates: `www.election.de`. OLS, respectively ordered probit regressions. The indicator variable conservative government equals one if a state's prime minister is from the Christian Democratic Union (CDU), and zero otherwise. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

[a]Dummy variable for late adopter state (as defined in Section 2.6.5), that is, first cohort affected graduating in 2011 or later.

[b]Year of implementation of the reform ranging from 1 (2001) to 7 (2007). Saxony and Thuringia are coded as category 0.

Table A.8: Effects of the Reform (Extended Individual Control Variables)

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Reform	-0.018	-0.101	0.081	0.110	0.325**	-0.171
	(0.150)	(0.154)	(0.110)	(0.138)	(0.143)	(0.110)
Female	0.043	-0.021	-0.015	-0.011	-0.001	-0.014
	(0.069)	(0.067)	(0.061)	(0.055)	(0.054)	(0.068)
Rural area	-0.049	-0.036	0.016	-0.015	0.115**	-0.036
	(0.058)	(0.099)	(0.076)	(0.062)	(0.045)	(0.058)
Non-intact family	0.116**	-0.075	0.155	-0.082	0.049	-0.113*
	(0.043)	(0.092)	(0.093)	(0.061)	(0.072)	(0.054)
Working-class father	-0.285***	0.040	0.010	-0.060	0.093	-0.006
	(0.084)	(0.089)	(0.088)	(0.087)	(0.094)	(0.064)
High parental educ.	0.016	-0.218***	-0.127	-0.054	0.003	0.038
	(0.052)	(0.048)	(0.076)	(0.061)	(0.041)	(0.054)
Working mother	0.075	-0.062	0.139**	0.012	-0.151**	0.005
	(0.061)	(0.064)	(0.061)	(0.043)	(0.054)	(0.048)
Migration background	0.226**	-0.011	-0.084	-0.080	-0.003	0.015
	(0.087)	(0.088)	(0.071)	(0.104)	(0.099)	(0.061)
Low-perform. student	-0.024	-0.032	-0.265**	0.091	0.097*	-0.188
	(0.129)	(0.086)	(0.110)	(0.113)	(0.055)	(0.115)
Only child	0.049	-0.039	0.083	-0.053	0.038	-0.016
	(0.095)	(0.074)	(0.052)	(0.076)	(0.090)	(0.079)
Oldest child	0.063	0.100**	0.063	-0.050	0.005	-0.028
	(0.040)	(0.044)	(0.040)	(0.056)	(0.075)	(0.053)
R^2	0.060	0.068	0.070	0.052	0.074	0.078
Observations	1263	1263	1263	1263	1263	1168

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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.

Table A.9: Effects of the Reform (Controlling for Parental Personality)

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Reform	-0.065	-0.120	0.070	0.121	0.395***	-0.205
	(0.177)	(0.154)	(0.142)	(0.153)	(0.121)	(0.150)
Open. (Mother)	0.261***					
	(0.046)					
Open. (Father)	0.143***					
	(0.037)					
Consc. (Mother)		0.135***				
		(0.028)				
Consc. (Father)		0.176***				
		(0.032)				
Extrav. (Mother)			0.175***			
			(0.049)			
Extrav. (Father)			0.126***			
			(0.033)			
Agree. (Mother)				0.172***		
				(0.048)		
Agree. (Father)				0.095***		
				(0.028)		
Neurot. (Mother)					0.238***	
					(0.041)	
Neurot. (Father)					0.082*	
					(0.040)	
LoC (Mother)						0.177***
						(0.043)
LoC (Father)						0.141*
						(0.069)
R^2	0.117	0.100	0.093	0.084	0.127	0.130
Observations	1055	1058	1057	1057	1057	928

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, and a constant are included. The respective personality traits of the parents are age-standardized and averaged over all available observations, 3 at maximum, to reduce measurement error. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.10: Sensitivity Analyses 2.6.1 - 2.6.2: Age Effects and Measurement Issues

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Outcome Variables: Age-free residuals of Personality Traits						
Reform	-0.033	-0.093	0.044	0.120	0.417**	-0.098
	(0.160)	(0.192)	(0.153)	(0.106)	(0.168)	(0.086)
Observations	1316	1316	1316	1316	1316	1202
Heterogeneous Effects by time of the interview[a]						
Reform	-0.045	-0.061	0.114	0.063	0.307	-0.083
	(0.128)	(0.189)	(0.117)	(0.143)	(0.186)	(0.114)
Reform*Quarter 2	0.027	-0.084	-0.215**	0.190	0.167	-0.150
	(0.104)	(0.116)	(0.090)	(0.157)	(0.155)	(0.127)
Reform*Quarter 3	0.060	0.108	0.114	-0.028	-0.145	-0.073
	(0.160)	(0.242)	(0.280)	(0.372)	(0.275)	(0.214)
Reform*Quarter 4	-0.810*	0.187	-0.408	0.160	0.556	0.423**
	(0.379)	(1.194)	(0.899)	(0.764)	(0.779)	(0.153)
Observations	1316	1316	1316	1316	1316	1202
Including dummy for having graduated as further control						
Reform	-0.029	-0.090	0.038	0.133	0.371**	-0.120
	(0.146)	(0.180)	(0.125)	(0.121)	(0.148)	(0.108)
High school diploma	-0.307	0.127	-0.209	0.179	-0.088	-0.055
	(0.218)	(0.163)	(0.212)	(0.140)	(0.187)	(0.183)
Observations	1316	1316	1316	1316	1316	1202
Subsample: Youth Questionnaire						
Reform	0.003	0.107	0.104	0.337**	0.228	-0.209
	(0.210)	(0.256)	(0.186)	(0.128)	(0.186)	(0.180)
Observations	729	729	729	729	729	712

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. In addition to the displayed coefficients, female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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]In addition to their interaction with $REFORM$, quarter dummies for the time of the interview are included in this estimation.

Table A.11: Sensitivity Analyses 2.6.4: Estimation Model

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Wild cluster bootstrapped standard errors[a,c]						
Reform	-0.034	-0.088	0.035	0.135	0.370***	-0.122
	(0.134)	(0.177)	(0.121)	(0.117)	(0.136)	(0.101)
Observations	1316	1316	1316	1316	1316	1202
Ordered Probit Estimation[b]						
Reform	-0.074	-0.107	0.032	0.114	0.397***	-0.124
	(0.137)	(0.179)	(0.119)	(0.115)	(0.148)	(0.114)
Observations	1316	1316	1316	1316	1316	1202
Accounting for the problem of multiple hypotheses testing[a]						
	Outcome Variables: Overall Measure of Personality[d]					
	including Big Five			including Big Five and LoC		
Reform	0.121*			0.079*		
	(0.057)			(0.041)		
Observations	1316			1202		

Notes: SOEPv30 waves 2005 to 2013. In addition to the displayed coefficients, female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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]OLS regressions.

[b]Ordered Probit estimation with the standardized personality measures as categorical outcomes.

[c]The reported standard errors are wild cluster bootstrapped following Cameron et al. (2008).

[d]Overall personality measure obtained following Anderson (2008) to avoid the problem of multiple hypotheses testing.

Table A.12: Sensitivity Analyses 2.6.5 - 2.6.7: Selectivity, Double Cohort, and Implementation Effects

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Subsample: Lateadopter states[a]						
Reform	0.089	0.074	0.159	0.050	0.284	0.219
	(0.169)	(0.187)	(0.254)	(0.245)	(0.153)	(0.224)
Observations	742	742	742	742	742	677
Subsample: States without comprehensive schools[b]						
Reform	0.010	-0.326	0.266*	0.185	0.383	-0.101
	(0.264)	(0.176)	(0.136)	(0.169)	(0.224)	(0.127)
Observations	709	709	709	709	709	653
Subsample: Exclusion of double graduating cohort						
Reform	0.054	-0.299*	0.060	-0.128	0.493***	-0.092
	(0.133)	(0.154)	(0.191)	(0.155)	(0.157)	(0.170)
Observations	1093	1093	1093	1093	1093	1017
Subsample: Exclusion of Saxony and Thuringia						
Reform	-0.058	-0.050	0.085	0.141	0.345*	-0.126
	(0.145)	(0.189)	(0.122)	(0.126)	(0.163)	(0.118)
Observations	1193	1193	1193	1193	1193	1087
Heterogeneous Effects by Time since Implementation of Reform[c]						
Reform	-0.038	-0.033	0.073	0.092	0.329*	-0.088
	(0.185)	(0.193)	(0.139)	(0.092)	(0.177)	(0.136)
Cohort=2	0.027	-0.113	0.206	-0.025	0.086	-0.082
	(0.192)	(0.146)	(0.204)	(0.100)	(0.070)	(0.112)
Cohort≥3	0.179	-0.135	0.360	-0.355***	0.084	0.074
	(0.241)	(0.179)	(0.210)	(0.114)	(0.203)	(0.139)
Observations	1193	1193	1193	1193	1193	1087

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Only the coefficients of *REFORM* and the respective number of observations are presented. Female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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]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.

[b]The states where comprehensive schools typically do not exist are Baden-Wuerttemberg, Bavaria, Lower Saxony, Mecklenburg-West Pomerania, Saxony, Saxony-Anhalt, and Thuringia.

[c]Next to *REFORM*, dummies for the second cohort affected by the reform and for cohorts three and higher are included in the estimation and presented. Saxony and Thuringia are naturally excluded from the sample in this estimation.

Table A.13: Sensitivity Analyses 2.6.8: Other Institutional Changes

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Subsample: States with standardized exit examinations established[a]						
Reform	-0.014	-0.211	0.190	0.214	0.483*	-0.101
	(0.294)	(0.151)	(0.131)	(0.181)	(0.205)	(0.112)
Observations	573	573	573	573	573	535

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Only the coefficients of *REFORM* and the respective number of observations are presented. Female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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]These states with central exit examinations long in place are Baden-Wuerttemberg, Bavaria, Mecklenburg-West Pomerania, Saarland, Saxony, Saxony-Anhalt and Thuringia.

Table A.14: Sensitivity Analyses 2.6.9: Placebo Estimation

	Outcome Variables: Personality Traits					
	Open.	Consc.	Extrav.	Agree.	Neurot.	LoC
Sample: Students from other school tracks						
Reform	0.146	0.182	-0.065	0.073	-0.127	-0.006
	(0.126)	(0.160)	(0.126)	(0.119)	(0.153)	(0.144)
R^2	0.041	0.084	0.030	0.037	0.025	0.066
Observations	1643	1643	1643	1643	1643	1458

Notes: SOEPv30 waves 2005 to 2013. OLS regressions. Female, age, age squared, state specific linear time trends, and a maximum set of state dummies, year of school entry dummies, dummies for the different SOEP samples, 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.

Appendix B

Appendix to Chapter 3

B.1 Samples and Variables

Table B.1: Sample Construction SOEP

Sample Description	Remaining observations
All respondents of the adolescent cognitive skills assessment in waves 2006 to 2013 with valid information on status of cognitive skill assessment test taken, federal state, year and month of birth	2,287
Drop students from school tracks other than high school or without information on track choice	900
Drop students from Rhineland-Palatinate	863
Drop students from Hesse entering high school in 2004 and 2005	849
Drop students who repeated one or more grades	787
Drop students who refused to take the cognitive skills test	748
Drop students who could not answer the test for technical reasons, i.e. no interviewer present	725
Drop students without valid information on background variables	724
Drop students without test scores on *all* dimensions of cognitive skills	723
Final Sample	723

Notes: SOEPv30 waves 2006 to 2013, youth questionnaires.

Table B.2: Sample Construction NEPS

Sample Description	Remaining observations
All respondents with valid information on year and month of birth	2,345
Drop students who repeated one or more grades	2,226
Drop students born in 1990 or 1999	2,224
Drop students without valid background information	2,128
Final Sample Maths	2,125
Final Sample Speed	2,128
Final Sample Reasoning	2,128

Notes: NEPS:BW:3.0.0 wave 2011/2012.

Table B.3: 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 B.4: 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.

B.2 Summary Statistics

Table B.5: 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 B.6: 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 B.7: 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 B.8: 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 B.9: 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.

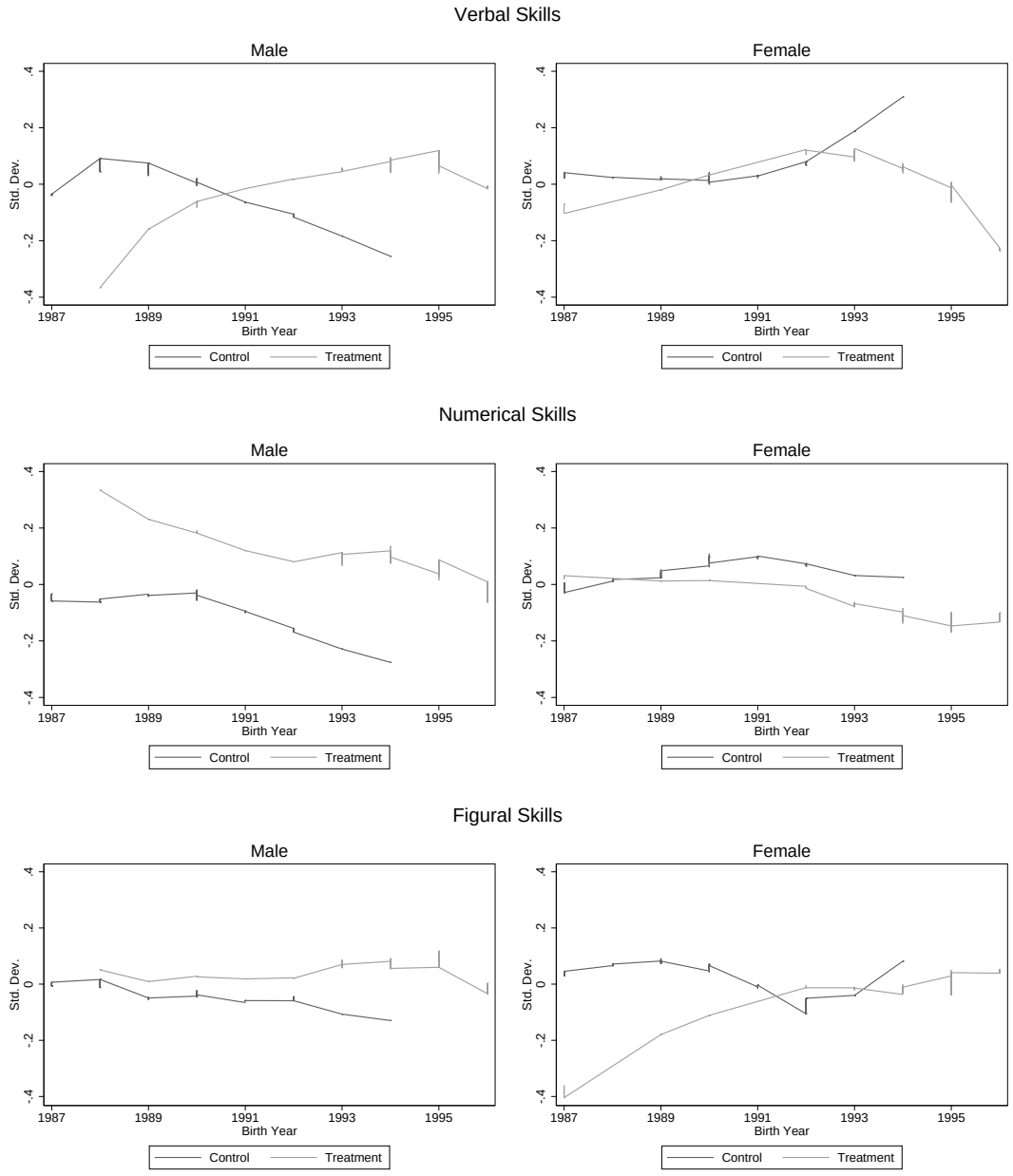
Table B.10: Summary Statistics of Individual Characteristics in NEPS

	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.

B.3 Figures

Figure B.1: 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.

B.4 Estimation Results on Instructional Time

Table B.11: Subsample of States with High School Starting in Grade 5

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Reform	0.097	0.368**	0.151
	(0.065)	(0.130)	(0.126)
Reform*Female	-0.006	-0.237**	-0.090
	(0.176)	(0.092)	(0.085)
R^2	0.114	0.087	0.117
Observations	637	637	637

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 B.12: Persistence of Effects

	Outcome Variables: Cognitive Skills in Adulthood	
	Crystallized[a]	Fluid[b]
Reform	0.277	0.454
	(0.497)	(0.626)
Reform*Female	-0.569	-0.590
	(0.576)	(0.768)
R^2	0.400	0.345
Observations	77	80

Notes: SOEPv30 waves 2006 to 2012. OLS regressions. Further, a maximum set of age dummies, 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.

[a]These are standardized measures of scores from a Multiple Choice-Vocabulary-Intelligence Test. See Lehrl et al. (1991) for details.

[b]These are standardized measures of scores from a 90 seconds Symbol-Digit-Test. See Lang et al. (2007) for details.

Table B.13: Sensitivity Analyses – Methodology

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Wild cluster bootstrapped standard errors[a]			
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[b]			
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[c]		
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.
[a]The reported standard errors are wild cluster bootstrapped following Cameron et al. (2008).
[b]This estimation further includes a linear time trend for each state.
[c]Overall cognitive skill measure obtained following Anderson (2008) to avoid the problem of multiple hypotheses testing.

Table B.14: 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[a]			
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.
[a]The estimation includes a dummy for each month of interview.

Table B.15: 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[a]			
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.
[a]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 B.16: Sensitivity Analyses – Selectivity

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Subsample: Late adopter states[a]			
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.
[a]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 B.17: 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 B.18: 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.

B.5 Estimation Results on Timing of Instruction

Table B.19: 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
Observations	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 B.20: 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[a]			
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
School Fixed Effects[a,b]			
Reform	0.022	-0.061	-0.070
	(0.047)	(0.067)	(0.048)
Reform	0.030	-0.094	-0.079
	(0.073)	(0.088)	(0.070)
Reform*Female	-0.015	0.059	0.015
	(0.081)	(0.093)	(0.089)
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.

[a]These estimations further control for individual characteristics including migration background, born January–June, high parental education, working-class father, working mother, and books at home.

[b]In addition, this estimation includes school fixed effects for the 48 different schools.

Table B.21: 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[a]			
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.
[a]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.

Appendix C

Appendix to Chapter 4

C.1 Sample and Variables

Table C.1: Sample Construction

Sample Description	Remaining observations
All available observations on MCS	150,925
Drop observations from East Germany via federal state *and* question where individuals lived in 1989	108,284
Drop observations of age younger than 50 and older than 85	49,087
Drop observations without valid information on schooling degree obtained	48,246
Drop observations without valid information on background variables	45,358
Reform Sample	
Keep only individuals born up to 7 years prior or post the cutoff birth year for being affected by the reform	19,185
Minimum Distance Sample	
Keep only individuals with place of birth (in West Germany) available	21,426

Notes: SOEPv31.1 waves 2002 to 2014.

Table C.2: SF-12 Questionnaire in SOEP

Question	**Answers**
How would you describe your current health?	Very good/ Good/ Satisfactory/ Poor/ Bad
When you have to climb several flights of stairs on foot, does your health limit you greatly, somewhat, or not at all?	Greatly/ Somewhat/ Not at all
And what about other demanding everyday activities, such as when you have to lift something heavy or do something requiring physical mobility: Does your health limit you greatly, somewhat, or not at all?	Greatly/ Somewhat/ Not at all
During the last four weeks, how often did you...	Always/ Often/ Sometimes/ Almost never/ Never
... feel down and gloomy?	
... feel calm and relaxed?	
... feel energetic?	
... have severe physical pain?	
... feel that due to physical health problems you achieved less than you wanted to at work or in everyday activities?	
... feel that due to physical health problems you were limited in some way at work or in everyday activities?	
... feel that due to mental health or emotional problems you achieved less than you wanted to at work or in everyday activities?	
... feel that due to mental health or emotional problems you carried out your work or everyday tasks less thoroughly than usual?	
... feel that due to physical or mental health problems you were limited socially, that is, in contact with friends, acquaintances, or relatives?	

Notes: MCS is calculated using factor analysis based on *all* items, including physical and mental health. The same holds true for PCS. Bold items denote mental health dimension.

Table C.3: Description of Variables

Variable	Description
Outcome Variables	
MCS	Standardized Mental Component Summary Score
PCS	Standardized Physical Component Summary Score
Depression	Dummy for having ever been diagnosed by a doctor with depression
Dementia	Dummy for having ever been diagnosed by a doctor with dementia
Sleep disorders	Dummy for having ever been diagnosed by a doctor with sleep disorders
Diagnoses	Dummy for having ever been diagnosed by a doctor with at least one of: depression, dementia, or sleep disorders
Intermediate degree or higher	Dummy for having at least an intermediate school leaving degree or higher
Dropout	Dummy for not having any school leaving degree
Control Variables	
yeduc	Years of schooling computed from secondary school degree
Reform	Dummy for being affected by the compulsory schooling reform
Mindist	Minimum spatial distance to university or university of applied sciences at age 19 (in deciles with respect to sample)
Age	Age (in years)
Female	Dummy for female
Non-German nationality	Dummy for non-German nationality
Rural area	Dummy for living in a rural area as classified by the Federal Institute for Research on Building, Urban Affairs and Spatial Development
High parental education	Dummy for at least one of an individual's parents having an upper secondary school degree or higher

Notes: SOEPv31.1 waves 2002 to 2014.

C.2 Summary Statistics

Table C.4: Summary Statistics of (non-standardized) Aggregate Health Measures

	Mean	Std. Dev.	Min.	Max.	No. of obs.
OLS Sample					
MCS	51.61	10.192	3.533	79.432	45358
PCS	45.489	10.231	10.698	76.421	45358
Reform Sample					
MCS	51.336	10.005	3.533	79.432	19185
PCS	47.46	9.808	11.147	75.456	19185
Minimum Distance Sample					
MCS	51.88	9.827	5.649	79.432	21426
PCS	46.604	9.843	12.59	76.421	21426

Notes: SOEPv31.1 waves 2002 to 2014, different samples. The distribution of the aggregate health measures prior to standardization is shown. Hence, the numbers correspond to the original scale provided in the data.

Table C.5: Means of Diagnoses in the Different Samples

	OLS Sample	Reform Sample	Min. Dist. Sample
Depression	0.068	0.083	0.069
Dementia	0.004	0.002	0.003
Sleep disturbances	0.120	0.126	0.113
Diagnoses	0.132	0.144	0.134
No. of observations[a]	20311	9605	12050

Notes: SOEPv31.1 waves 2009, 2011, and 2013, different samples.
[a]For sleep disturbances, the number of observations are lower with 13876, 6525, and 8938, as this variable is not observed in 2009.

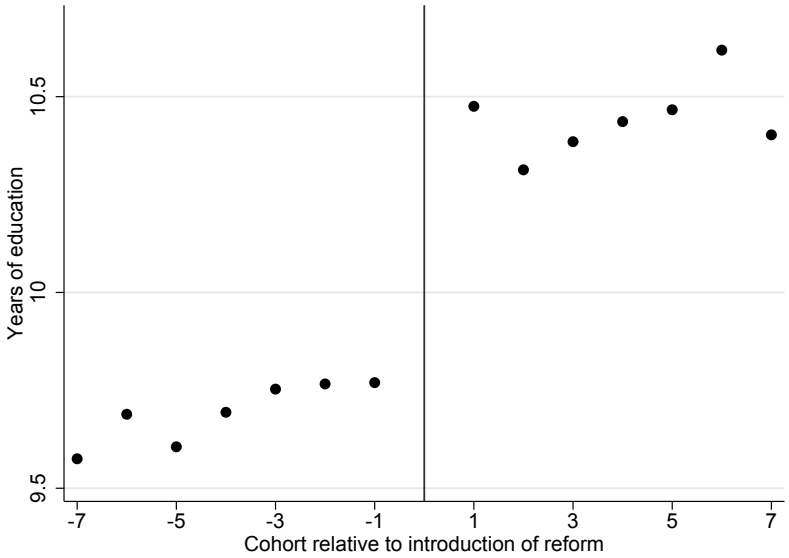
Table C.6: Means of Other Variables in the Different Samples

	OLS Sample	Reform Sample	Min. Dist. Sample
Instruments			
Reform		0.437	
Mindist (in km)			21.499
Control variables			
yeduc	9.756	10.019	9.912
Age	63.810	58.153	62.607
Female	0.514	0.513	0.507
Non-German nationality	0.061	0.070	0.003
Rural area	0.247	0.256	0.271
High parental education	0.124	0.129	0.126
No. of observations	45358	19185	21426

Notes: SOEPv31.1 waves 2002 to 2014, different samples.

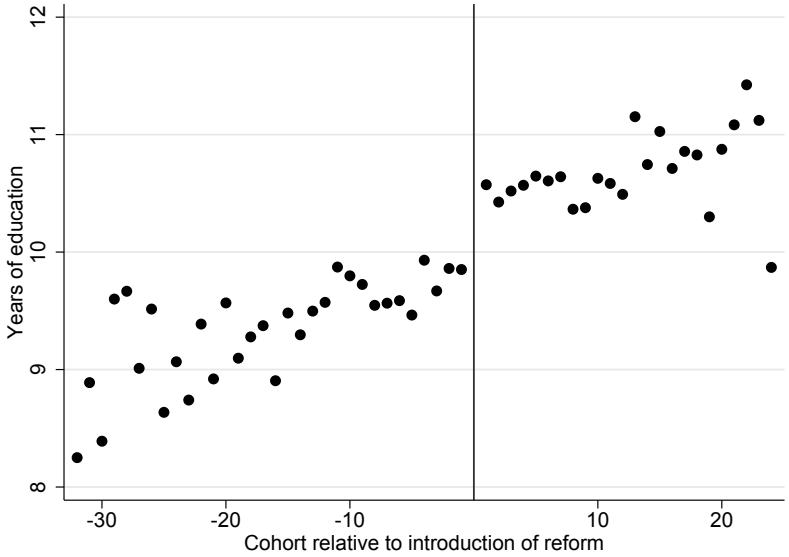
C.3 Figures

Figure C.1: Average Number of Years of Schooling by Birth Cohort



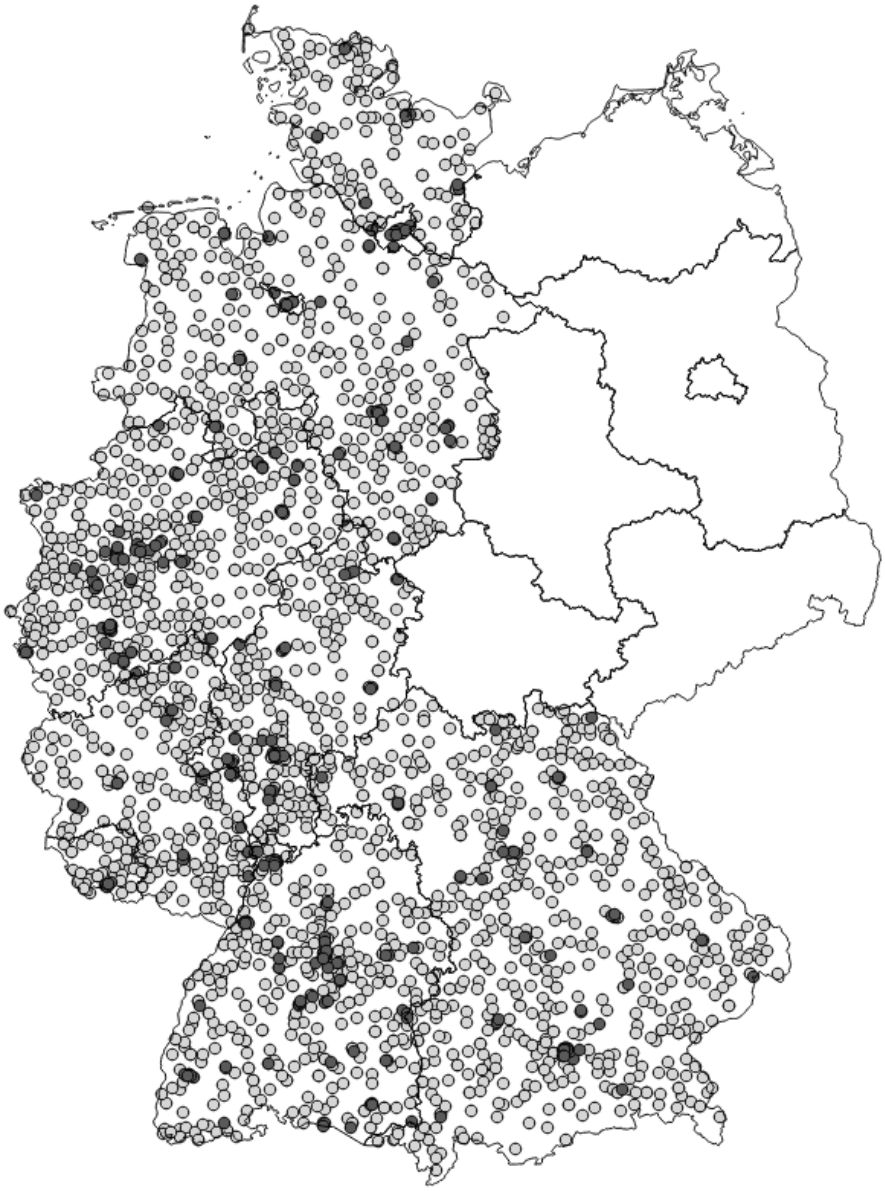
Own illustration, *Data Source:* SOEPv31.1 Reform Sample.

Figure C.2: Average Number of Years of Schooling by Birth Cohort



Own illustration, *Data Source:* SOEPv31.1 Minimum Distance Sample. Averages of cohorts with at least 20 observations are displayed, together with linear fitted lines.

Figure C.3: Geographical Location of Birth Places and Universities (West Germany)



Own illustration, *Data Sources:*
○ Places of birth: SOEP waves 2012-2014
● Universities (existent in 2016): Stiftung zur Förderung der Hochschulrektorenkonferenz

Figure C.4: Growth of Universities over Time (West Germany)

Own illustration, *Data Source:*
● Universities (existent in the respective year): Stiftung zur Förderung der Hochschulrektorenkonferenz

C.4 Estimation Results

Table C.7: Effects of the Reform on Track Choice and Success

	Outcome variables	
	Intermediate Degree or higher	Dropout
	(1)	(2)
Reform	0.039	-0.002
	(0.041)	(0.010)
R^2	0.107	0.075
No. of observations	19185	19185
No. of persons	5789	5789

Notes: SOEPv31.1 waves 2002 to 2014, Reform Sample. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. * p<0.1, ** p<0.05, *** p<0.01.

Table C.8: IV Estimations – Sensitivity Analyses: Sample

	Outcome variable: MCS			
	Reform Sample		Minimum Distance Sample	
	Weighted estimation[a] (1)	One obs. per person[b] (2)	Weighted estimation[a] (3)	One obs. per person[b] (4)
Second stage results				
yeduc	-0.196	-0.177	0.030	0.072
	(0.132)	(0.152)	(0.103)	(0.122)
R^2	-0.084	-0.054	0.048	0.059
No. of observations	19185	5789	21426	5250
No. of persons	5789		5250	
First stage results on yeduc				
Reform	0.530***	0.523***		
	(0.116)	(0.118)		
Mindist			-0.043***	-0.043***
			(0.009)	(0.009)
F-stat.	20.735	19.682	22.521	23.475

Notes: SOEPv31.1 waves 2002 to 2014. 2SLS estimations. Further, a maximum set of state dummies, year of birth dummies, survey year dummies, linear birth year state trends, and a constant are included. Individual characteristics controlled for include age squared, female, non-German nationality, rural area, and high parental education. Standard errors, reported in parentheses, are clustered at individual level. F-stat. denotes the Kleibergen-Paap Wald rk F statistic for weak identification in case of clustered standard errors. * p<0.1, ** p<0.05, *** p<0.01.

[a]Observations are weighted by the inverse frequency of the individual observed in the sample.

[b]For each individual the closest observation to age 67.5 is chosen. If there are two, the one at younger age is taken.