ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Arold, Benjamin W.

Research Report

School curricula, educational trajectories, and labor market outcomes

ifo Beiträge zur Wirtschaftsforschung, No. 99

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Arold, Benjamin W. (2022) : School curricula, educational trajectories, and labor market outcomes, ifo Beiträge zur Wirtschaftsforschung, No. 99, ISBN 978-3-95942-117-1, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, München

This Version is available at:
https://hdl.handle.net/10419/263222

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ifo
BEITRÄGE
zur Wirtschaftsforschung

99
2022

School Curricula, Educational Trajectories, and Labor Market Outcomes

Benjamin W. Arold

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

ifo
BEITRÄGE
zur Wirtschaftsforschung

99
2022

School Curricula, Educational Trajectories, and Labor Market Outcomes

Benjamin W. Arold

Herausgeber der Reihe: Clemens Fuest
Schriftleitung: Chang Woon Nam

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über
http://dnb.d-nb.de
abrufbar.

ISBN: 978-3-95942-117-1

Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten.
Ohne ausdrückliche Genehmigung des Verlags ist es auch nicht gestattet, dieses
Buch oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie)
oder auf andere Art zu vervielfältigen.
© ifo Institut, München 2022

Druck: ifo Institut, München

ifo Institut im Internet:
https://www.ifo.de

Preface

Benjamin W. Arold prepared this study while he was working at the Center for Economics of Education at the ifo Institute. The study was completed in March 2022 and accepted as doctoral thesis by the Department of Economics at the LMU Munich. It consists of four distinct empirical essays that address various aspects of how school curricula affect students in the classroom and beyond. Chapter 2 demonstrates that the abolishment of compulsory religious education in Germany does not only affect religiosity and gender role attitudes, but also affects the high-stakes choices of getting married, having children and participating in the labor market. Chapter 3 provides evidence that the content of science education standards in the US does not only affect related knowledge of students and adulthood attitudes, but also translates into high-stakes occupational choices. Chapter 4 exemplifies that school curricula reforms can also have unintended consequences by showing that a reform of education standards in math and English has negative spillovers on student achievement in science. Finally, Chapter 5 highlights the fundamental role teachers, the transmitters of educational content in school, play in the formation of human capital even in dimensions generally thought of as genetically determined.

Keywords: Religious Education, Religiosity, School Reforms, Evolution, Religion, Science Education, Education Standards, Human Capital, Attitudes, Beliefs, Occupational Choice, Common Core, Education Standards, Student Achievement, Education Policy, Polygenic Scores, School Resources, Skill Formation

JEL-No: I21, I24, I28, I29, J24, H75, P16, Z12

For Edrea

Acknowledgements

First and foremost, I would like to thank my advisor Ludger Woessmann for always supporting and believing in me, for providing invaluable feedback on numerous occasions, for having given me the academic freedom to develop and pursue own research interests, and for having created a warm and welcoming atmosphere at the research center which made me enjoy going to office and becoming a researcher. Your approach to combine excellent research with policy relevance has been inspirational, and it has been a pleasure and honor to be your co-author and learn from you.

Furthermore, I am very grateful to my second advisor Davide Cantoni who has supported me in countless sessions on research ideas, feedback on drafts, career planning, and for having written numerous letters of reference for me since my days as an undergraduate student. I would also to thank Andreas Peichl for very helpful feedback, for making me feel welcome whenever visiting his research group, and for joining my committee.

I am deeply endebted to Larissa Zierow. I could not have asked for a more supportive, engaging, and brilliant mentor and co-author since my first day at the ifo Institute. I have cherished our countless meetings and lunch walks, and much appreciated that your door has been always open for guidance and advice. I also thank my other co-authors Paul Hufe, Marc Stoeckli, M. Danish Shakeel, Vera Freundl, and Katia Werkmeister for great and reliable co-authorship.

I thank my colleagues at the ifo Center for the Economics of Education for wonderful company, and insightful discussions. Special thanks to Ulrike Baldi-Cohrs and Franziska Binder for providing outstanding support in many areas far beyond administrative tasks.

During the academic year 2019/2020, I had the incredible opportunity to visit Harvard University. I am most grateful to Paul Peterson, Marty West, Antonio Wendland and the whole team of PEPG (Program on Education Policy and Government) who took me in as one of their own. I would also like to thank the DAAD for funding my stay at Harvard.

I am very grateful to my peers and friends in academia and beyond. In particular, I would like to thank Lavinia Kinne and Lisa Simon for having been amazing office neighbors in the north wing of the first floor, and Elisabeth Grewenig for invaluable help and countless laughs.

Last, but not least, I would like to express my deepest gratitude to my parents and family. I am incredibly thankful for your constant support and faith in me.

Benjamin Wilhelm Arold
July 2022

School Curricula, Educational Trajectories, and Labor Market Outcomes

Inaugural-Dissertation
Zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)

eingereicht an der
Ludwig-Maximilians-Universität München
2022

vorgelegt von

Benjamin W. Arold

Referent: Prof. Dr. Ludger Wößmann
Korreferent: Prof. Davide Cantoni, Ph.D.
Promotionsabschlussberatung: 20.07.2022

Datum der mündlichen Prüfung:	06.07.2022
Namen der Berichterstatter:	Prof. Dr. Ludger Wößmann
	Prof. Davide Cantoni, Ph.D.
	Prof. Dr. Andreas Peichl

Contents

Preface I

Acknowledgements V

List of Figures XIII

List of Tables XV

1 General Introduction 1
- 1.1 Education and Human Capital . . . 1
- 1.2 Education Policy and School Curricula Reforms . . . 2
- 1.3 Empirical Methods for Causal Inference . . . 3
- 1.4 Data . . . 4
- 1.5 Chapter Overview . . . 5

2 Can Schools Change Religious Attitudes? 9
- 2.1 Introduction . . . 9
- 2.2 Institutional Background: Reforms Abolishing Compulsory Religious Education in Germany . . . 12
  - 2.2.1 Historical Background . . . 12
  - 2.2.2 Post-War Situation . . . 13
  - 2.2.3 The Reforms . . . 13
- 2.3 Empirical Model . . . 16
- 2.4 Data . . . 19
- 2.5 The Effect of Abolishing Compulsory Religious Education on Religiosity . . . 20
- 2.6 Effects on Family and Labor-Market Outcomes . . . 22
- 2.7 Robustness . . . 23
  - 2.7.1 Placebo test: Effects on Non-Religious School Outcomes . . . 23
  - 2.7.2 Border Specification with County-Pair Fixed Effects . . . 23
  - 2.7.3 Additional Robustness Analyses . . . 24
  - 2.7.4 Tests of the Two-Way Fixed Effects Estimator . . . 24
- 2.8 Conclusion . . . 25
- Figures and Tables . . . 27
- Appendix . . . 39

3 Evolution vs. Creationism in the Classroom 77
- 3.1 Introduction . . . 77

3.2 Institutional Background . . . 81
3.2.1 The Battle for Teaching Evolution in US Public Schools . . . 81
3.2.2 US State Science Standards . . . 82
3.2.3 Reform Examples from Florida and Texas . . . 83
3.2.4 The Adoption Process of Reforms of Science Standards . . . 84
3.2.5 The Implementation of Reforms of Science Standards . . . 85
3.3 Data . . . 85
3.3.1 Coding of Reforms of Science Standards . . . 85
3.3.2 Micro Data . . . 87
NAEP: Evolution Knowledge in School . . . 87
GSS: Evolution Approval in Adulthood . . . 88
ACS: Occupational Choice . . . 89
3.4 Identification Strategy . . . 89
3.5 Results . . . 92
3.5.1 Evolution Knowledge in School . . . 93
3.5.2 Evolution Approval in Adulthood . . . 94
3.5.3 Occupational Choice . . . 95
3.6 Robustness . . . 96
3.6.1 Event-Study Graphs . . . 96
3.6.2 Placebo Tests . . . 97
3.6.3 Time-Varying Treatment Effects . . . 99
3.6.4 Further Robustness Checks . . . 100
3.7 Conclusion . . . 101
Figures and Tables . . . 104
Appendix . . . 123

4 Unintended Effects of the Common Core State Standards 151
4.1 Introduction . . . 151
4.2 Institutional Background . . . 155
4.2.1 Background and Data on the Adoption of the CCSS . . . 155
4.2.2 Background and Data on the Implementation of the CCSS . . . 156
4.3 Identification Strategy . . . 156
4.4 Data . . . 160
4.4.1 Student Achievement Data . . . 160
4.4.2 Descriptive Statistics . . . 162
4.5 Results . . . 162
4.5.1 Main Results . . . 162
4.5.2 Subgroup Analysis . . . 163
4.6 Robustness . . . 164
4.6.1 Robustness Tests on Parallel Trends Assumption . . . 164
4.6.2 Robustness Tests on Time-Varying Treatment Effects . . . 166
4.6.3 Robustness Tests on Treatment Definition . . . 166

4.6.4 Further Specification Checks . . . 167
4.7 Mechanisms . . . 168
4.8 Conclusion . . . 169
Figures and Tables . . . 171
Appendix . . . 179

5 Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments 211
5.1 Introduction . . . 211
5.2 Measuring Genetic Endowment . . . 215
5.3 Empirical Strategy . . . 217
5.3.1 Empirical Model . . . 217
5.3.2 Identification . . . 218
5.4 Data . . . 220
5.5 Results . . . 224
5.5.1 The Association of Educational Attainment with Genetic Endowments and School Investments . . . 224
5.5.2 The Interplay of Genetic Endowments and School Investments in the Production of Educational Attainment . . . 226
5.5.3 Robustness Analysis . . . 228
5.5.4 Mechanisms . . . 231
5.6 Conclusion . . . 233
Figures and Tables . . . 235
Appendix . . . 246

Bibliography 265

List of Figures

Figure 2.1: The effect of abolishing compulsory religious education on religiosity: Non-parametric event-study estimates 28
Figure A2.1: Religious education reforms in West German states 46
Figure A2.2: Non-parametric event-study estimates of effect on personal prayer . . 47
Figure A2.3: Non-parametric event-study estimates of effect on church-going . . . 48
Figure A2.4: Non-parametric event-study estimates of effect on religious affiliation 49
Figure A2.5: Counties in the border specification 50

Figure 3.1: US map of evolution score difference between 2000 and 2009 116
Figure 3.2: Effect of evolution coverage in Science Standards on evolution knowledge in school, by subgroups 117
Figure 3.3: Effect of evolution coverage in Science Standards on evolution approval in adulthood, by subgroups 118
Figure 3.4: Effect of evolution coverage in Science Standards on probability to work in life sciences, by subfields of life sciences 119
Figure 3.5: Event-study graphs: Evolution knowledge in school 120
Figure 3.6: Event-study graphs: Evolution approval in adulthood 121
Figure 3.7: Event-study graphs: Occupational choice 122
Figure A3.1: Teachers' focus on State Science Education Standards when teaching evolution (Teacher Survey) 148
Figure A3.2: US map of evolution scores 149
Figure A3.3: Two NAEP sample questions on evolution knowledge 150

Figure 4.1: CCSS adoption map 172
Figure 4.2: Event-study graph: Non-targeted subjects 173
Figure A4.1: Event-study graph: Targeted subjects (NAEP) 196
Figure A4.2: Event-study graph: Targeted subjects (SEDA) 197

Figure 5.1: Rotated loadings on factors for school characteristics 236
Figure 5.2: Association of educational attainment with PGS^{EA}, $I_{Quality}$, and $I_{Quantity}$. 237
Figure 5.3: PGS^{EA} distribution by $I_{Quality}$ and $I_{Quantity}$ 238
Figure 5.4: Association of PGS^{EA} with years of education by $I_{Quality}$ 239
Figure 5.5: Association of PGS^{EA} and school/family environments with degree attainment 240
Figure A5.1: Top 3 school expenditure categories (in % of total) 253
Figure A5.2: Sensitivity to unobserved confounders 254

List of Tables

Table 2.1: The rollout of abolishing compulsory religious education: Timing of treatment and governing parties . . . 29
Table 2.2: Descriptive statistics . . . 30
Table 2.3: Effects of abolishing compulsory religious education on religiosity and religious actions . . . 32
Table 2.4: Time-varying treatment effects on religious outcomes . . . 33
Table 2.5: Heterogeneous treatment effects on religious outcomes . . . 34
Table 2.6: Effects on ethical-value and political-value outcomes . . . 35
Table 2.7: Effects on family and labor-market outcomes . . . 36
Table 2.8: Effects on educational outcomes . . . 37
Table 2.9: Effects on religious outcomes: Border specification with county-pair fixed effects . . . 38
Table A2.1: Curricula before and after the reform: The case of Bavaria . . . 51
Table A2.2: Outcome measures derived from the three datasets . . . 53
Table A2.3: Effects on religious outcomes: Non-parametric estimation . . . 57
Table A2.4: Effect on religiosity: Dummy coding of religiosity . . . 58
Table A2.5: Effect on religiosity: Controls for other school reforms . . . 59
Table A2.6: Effect on personal prayer: Controls for other school reforms . . . 60
Table A2.7: Effect on church-going: Controls for other school reforms . . . 61
Table A2.8: Effect on religious affiliation: Controls for other school reforms . . . 62
Table A2.9: Effects on religious outcomes: Dosage treatment . . . 63
Table A2.10: Effects on religious outcomes: Reform timing coded at secondary school entry . . . 64
Table A2.11: Effects on religious outcomes: Excluding early reforming states . . . 65
Table A2.12: Effects on religious outcomes: Only NEPS data . . . 66
Table A2.13: Effects on religious outcomes: Only ALLBUS data . . . 67
Table A2.14: Effects on religious outcomes: Only SOEP data . . . 68
Table A2.15: Effects on religious outcomes: Callaway and Sant'Anna estimator . . . 69
Table A2.16: Effects on religious outcomes: Excluding Berlin and Bremen . . . 70
Table A2.17: Goodman-Bacon decomposition of effect on religiosity . . . 71
Table A2.18: Goodman-Bacon decomposition of effect on personal prayer . . . 72
Table A2.19: Goodman-Bacon decomposition of effect on church-going . . . 73
Table A2.20: Goodman-Bacon decomposition of effect on religious affiliation . . . 74
Table A2.21: Effects on religious outcomes: No control variables . . . 75

Table 3.1: Effect of evolution coverage in Science Standards on evolution knowledge in school . . . 105

Table 3.2: Effect of evolution coverage in Science Standards on evolution approval in adulthood . . . 106
Table 3.3: Effect of evolution coverage in Science Standards on probability to work in life sciences . . . 107
Table 3.4: Placebo tests: Effect of evolution coverage in Science Standards on knowledge about non-evolution scientific topics . . . 108
Table 3.5: Placebo tests: Effect of evolution coverage in Science Standards on evolution knowledge in private schools . . . 109
Table 3.6: Placebo tests: Effect of evolution coverage in Science Standards on non-evolution scientific topics . . . 110
Table 3.7: Placebo tests: Effect of evolution coverage in Science Standards on religious outcomes . . . 111
Table 3.8: Placebo tests: Effect of evolution coverage in Science Standards on political outcomes . . . 112
Table 3.9: Effect of evolution coverage in Science Standards on probability to work in broader scientific occupational subfields . . . 113
Table 3.10: Placebo test: Effect of evolution coverage in Science Standards on probability to work in non-scientific occupational fields . . . 114
Table A3.1: Evolution scores and reform year, by state . . . 128
Table A3.2: Effect of evolution coverage in Science Standards on probability to work in life sciences, by subgroups . . . 129
Table A3.3: Effect of evolution coverage in Science Standards on evolution knowledge in school: CS estimator . . . 130
Table A3.4: Effect of evolution coverage in Science Standards on evolution approval in adulthood: CS estimator . . . 131
Table A3.5: Effect of evolution coverage in Science Standards on probability to work in life sciences: CS estimator . . . 132
Table A3.6: Effect of evolution coverage in Science Standards on evolution knowledge, further robustness checks . . . 133
Table A3.7: Effect of evolution coverage in Science Standards on evolution approval in adulthood, further robustness checks . . . 134
Table A3.8: Effect of evolution coverage in Science Standards on probability to work in life sciences, further robustness checks . . . 135
Table A3.9: Effect of evolution coverage in Science Standards on evolution knowledge in school, by evolution score indicator variables . . . 136
Table A3.10: Effect of evolution coverage in Science Standards on evolution approval in adulthood, by evolution score indicator variables . . . 137
Table A3.11: Effect of evolution coverage in Science Standards on probability to work in life sciences, by evolution score indicator variables . . . 138
Table A3.12: Correlation coefficients of knowledge about different scientific areas . . . 139
Table A3.13: Descriptive statistics of NAEP data . . . 140
Table A3.14: Correlation coefficients of approval to different scientific areas . . . 141

Table A3.15: Correlation coefficients of evolution approval and religious outcomes 142
Table A3.16: Correlation coefficients of evolution approval and political outcomes . 143
Table A3.17: Descriptive statistics of GSS data 144
Table A3.18: Descriptive statistics of ACS data 146

Table 4.1: Effect of CCSS exposure on student achievement in non-targeted subjects 174
Table 4.2: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by student characteristics 175
Table 4.3: Effect of CCSS exposure on student achievement in non-targeted subjects, econometric robustness 176
Table 4.4: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness using different definitions of treatment implementation . . 177
Table 4.5: Effect of CCSS exposure on instructional focus in non-targeted subjects 178
Table A4.1: List of grades of NAEP tests for non-targeted subjects 180
Table A4.2: Descriptive statistics 181
Table A4.3: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by subjects 183
Table A4.4: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by grades 184
Table A4.5: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by quartiles of states' student achievement before 2010 . . 185
Table A4.6: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for public schooling policies . . . 186
Table A4.7: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for private schooling policies . . . 187
Table A4.8: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for homeschooling policies 188
Table A4.9: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for compulsory schooling policies 189
Table A4.10: Effect of CCSS exposure on student achievement in non-targeted subjects, further robustness checks 190
Table A4.11: Effect of CCSS exposure on instruction time in non-targeted subjects, subgroups by teacher characteristics 191
Table A4.12: Effect of CCSS exposure on instructional resources in non-targeted subjects, subgroups by teacher characteristics 192
Table A4.13: Effect of CCSS exposure on the quality of teacher-student interactions (setting goals) in non-targeted subjects, subgroups by teacher characteristics 193
Table A4.14: Effect of CCSS exposure on the quality of teacher-student interactions (discussing goals) in non-targeted subjects, subgroups by teacher characteristics 194

Table A4.15: Effect of CCSS exposure on student achievement in targeted subjects, econometric robustness 198
Table A4.16: Effect of CCSS exposure on student achievement in targeted subjects using different definitions of treatment implementation 199
Table A4.17: List of grades of NAEP tests for targeted subjects 200
Table A4.18: State-mandated tests, by state and grade group from 2010 onwards (based on own research) 202
Table A4.19: Treatment codings by state 208

Table 5.1: Summary statistics 241
Table 5.2: Association of PGS^{EA} and school environments with years of education 242
Table 5.3: Robustness to additional school characteristics 243
Table 5.4: Robustness to family environments 244
Table 5.5: Association of PGS^{EA} and school environments with skill measures . . 245
Table A5.1: Sample representativeness 248
Table A5.2: Association of PGS and years of education: Comparing between-family and within-family models 249
Table A5.3: Tests for equality of PGS^{EA} distributions 250
Table A5.4: Alternative polygenic scores 251
Table A5.5: Robustness to sample composition 252
Table A5.6: Summary statistics (outcomes) 258
Table A5.7: Summary statistics (variables of interest) 259
Table A5.8: Summary statistics (controls) 263

1 Introduction

1.1 Education and Human Capital

The importance of education for personal and societal development has been acknowledged for a long time. In ancient Greece, Plato stated: "If a man neglects education, he walks lame to the end of his life". It has also long been acknowledged that, despite its potential, educational investments are not free of costs. The quote of Aristotle that "the roots of education are bitter, but the fruit is sweet" illustrates the intertemporal trade-off that underlies educational investment decisions. Modern economic theory has formalized this cost-benefit rationale in the so-called human capital theory (Mincer, 1958; Schultz, 1961; Becker, 1962). Becker (1962) argues that investments in education and training increase productivity levels of individuals and thereby improve their labor market outcomes such as wages and unemployment risks. At the same time, investments in education are costly. Individuals have to bear direct costs of education such as tuition fees, and indirect costs in the form of opportunity costs due to foregone time. Rational individuals maximize their utility by investing into education to the point in which the expected benefits equal the expected costs.

At the macroeconomic level, economic output can be modeled as a function of capital and labor inputs (Solow, 1956). In augmented neoclassical growth models, education can increase individual productivity and, in the aggregate, human capital (Mankiw et al., 1992). Another way to formalize the beneficial effects of educational investments for economic output comes from endogenous growth models. Here, education increases the innovative capacity of an economy, generating new technologies and leading to economic growth (Romer, 1990; Aghion and Howitt, 1998).

Complementing the theoretical literature, empirical studies in the economics of education have also demonstrated positive effects of educational expansions on individual and societal development. Individuals' increases in educational attainment have been consistently associated with wage increases (Card, 1999; Heckman et al., 2006). At the macroeconomic level, educational achievement, particularly in math and science, has been repeatedly linked with economic growth (Hanushek and Woessmann, 2008, 2012). Hence, understanding which determinants affect educational outcomes is of special interest to policy makers and researchers alike (Woessmann, 2016).

Education production functions link educational outcomes to inputs including family factors, school resources, and institutional elements of education systems (Hanushek, 1986; Woessmann, 2016). Much of the recent empirical literature in the economics of education has studied causal relationships between educational outcomes and their determinants. While family

background certainly plays an important role in the formation of human capital, isolating causal effects of educational resources and institutional determinants has been a primary focus, not least because of its immediate policy relevance. Examples of studies on the effects of school resources include analyses of the effects of school finances (Jackson et al., 2016; Lafortune et al., 2018; Jackson et al., 2021), class size (Angrist and Lavy, 1999; Krueger, 1999; Angrist et al., 2019), teacher quality (Chetty et al., 2014b; Jackson, 2018; Bacher-Hicks et al., 2019), and instruction time (Lavy, 2015; Rivkin and Schiman, 2015; Wedel, 2021). Studies on the institutional structure of the school system include topics such as central exit exams (Jürges et al., 2005; Schwerdt and Woessmann, 2017), school autonomy (Hanushek et al., 2013; Hahn et al., 2018), school accountability (Rouse et al., 2013; Feng et al., 2018; Bergbauer et al., 2021), and tracking (Hanushek and Woessmann, 2006; Matthewes, 2020; Grewenig, 2021).

1.2 Education Policy and School Curricula Reforms

While many features of education systems have received considerable attention in the literature, studies on the effects of school curricula are comparatively rare, arguably because associations of school curricula with outcomes are hard to quantify and exogenize. Cantoni and Yuchtman (2013, p. 243) present a theoretical model of how a governmental setting of educational content and individual education decisions interact. They conclude that "further study of educational content and educational institutions as crucial determinants of human capital in a society, and thus growth, appears to us to be an important, under-explored area of study". This call was echoed by Cantoni et al. (2017, p. 386) stating that "the choices that elites make regarding educational content deserve further study". This dissertation addresses these calls by showing how educational content can affect human capital measures such as student achievement and wages. Beyond typical human capital measures, effects of educational content on attitudinal outcomes as well as family and career decisions are studied as well.

The small but growing quasi-experimental literature on the effects of school curricula has generated first insights about how students' knowledge can be influenced by educational content. These papers are identified from variation in topic-specific instruction time (Machin and McNally, 2008; Cortes and Goodman, 2014), minimum high school course requirements (Goodman, 2019), and advanced placement courses (Conger et al., 2021). Beyond students' knowledge, attitudinal outcomes can also be affected as demonstrated for identity (Clots-Figueras and Masella, 2013), political and economic attitudes (Cantoni et al., 2017), cultural civic values (Bandiera et al., 2019), and religiosity (Bazzi et al., 2020).

The main contribution of this dissertation is to present empirical evidence that school curricula affect high-stakes choices of exposed students in adulthood. Chapter 2 demonstrates that the abolishment of compulsory religious education in Germany does not only affect religiosity and gender role attitudes, but also affects the high-stakes choices of getting married, having children and participating in the labor market. Chapter 3 provides evidence that the content

of science education standards in the US does not only affect related knowledge of students and adulthood attitudes, but also translates into high-stakes occupational choices. Chapter 4 exemplifies that school curricula reforms can also have unintended consequences by showing that a reform of education standards in math and English has negative spillovers on student achievement in science. Finally, Chapter 5 highlights the fundamental role teachers, the transmitters of educational content in school, play in the formation of human capital even in dimensions generally thought of as genetically determined.

1.3 Empirical Methods for Causal Inference

This dissertation aims to estimate the causal effect of school curricula on educational, labor market and further outcomes of exposed students. Simple correlations between school curricula and outcomes of interest likely do not yield causal effects. In general, school curricula are not randomly adopted but reflect the content educational policy makers deem relevant and appropriate for their students. For example, the religious attitudes of the population and policy makers may influence the religious content that is included in school curricula, which in turn may affect the religious attitudes of exposed students. Similarly, the academic rigor of an education standard may depend on the skill levels of current students, and in turn affect their achievement. In the context of the research questions about school curricula addressed in this dissertation, one can easily think of a multitude of factors along educational, economic, political, scientific, or religious dimensions that bias simple correlations of school curricula and outcomes away from the true effect of school curricula on outcomes.

The causal effect of school curricula on outcomes could be cleanly estimated if different school curricula were randomly assigned to students with everything else being held constant. Under such ideal experimental conditions, individuals in the treatment and control groups would be comparable in terms of observable and unobservable characteristics if the sample size is sufficiently large. Hence, any difference in outcomes could be causally attributed to the treatment (Angrist and Pischke, 2009; Schwerdt and Woessmann, 2020). However, such an experiment about school curricula appears unfeasible for practical and ethical reasons. Hence, the chapters of this dissertation exploit natural experiments that mimic the random assignment of an experiment. Natural experiments are observational studies in which exposure to treatment and control conditions is determined by "nature", i.e. by factors that are not manipulated by an experimenter but still orthogonal to characteristics affecting the outcome in question.

In chapters 2 to 4, I primarily exploit the staggered adoption of school curricula reforms across states and over time within a given country. Accounting for fixed effects for states and cohorts, two-way fixed effects models can provide plausibly exogenous variation in individuals' exposure to school curricula. Generally speaking, school curriculum effects in these models are identified from differences in outcomes between cohorts within the same state that were

and were not subject to a given school curriculum, relative to the differences between the same cohorts in other states that did not have school curriculum reforms at the same time.

The recent econometric literature on two-way fixed effects models has emphasized that an unbiased and consistent estimation of reform effects requires, in addition to the absence of confounding shocks, homogeneity in treatment effects (Chaisemartin and D'Haultfœuille, 2020; Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Roth and Sant'Anna, 2021; Sun and Abraham, 2021). In settings with staggered treatment timing like the ones presented in chapters 2 to 4, time-varying treatment effects can bias results away from the true effect if already-treated students act as controls for later-treated students (negative weighting). I address this concern in different ways depending on the chapter and setting therein, for example through robustness checks in which those 2x2 difference-in-differences comparisons in which already-treated students act as controls are excluded from the sample (Callaway and Sant'Anna, 2021).

To assess the validity and robustness of results further, I also conduct a wide range of supplementary econometric analyses. Depending on the chapter, these analyses include event-study models, triple difference specifications, placebo tests, and subsample analyses of areas that are geographically close or governed by narrowly elected governors, among others. In chapter 5, the relevant findings about the associations of genetic endowments with educational outcomes are replicated in sibling samples to control for genetic nurture effects, and in bounding exercises following Oster (2019) to account for selection effects.

1.4 Data

To run the econometric models, this dissertation combines various data sources. In chapters 2 to 4, legal data about school curricula is merged with large-scale microeconometric data containing individual-level information about exposed students. The legal data typically consist of evaluations of education laws and standards. They comprise state-year level data about the compulsory nature of school subjects in chapter 2 (Helbig and Nikolai, 2015), the topical content of education standards in chapter 3 (Lerner, 2000b; Mead and Mates, 2009), and the adoption of a certain type of education standards in chapter 4 (Achieve Inc., 2013; CCSSI, 2022). Importantly, the legal data have to measure the variables in question at multiple points in time for a given state to allow for identification of effects within states over time.

Each chapter of the dissertation makes use of large-scale survey data, either for the main analyses, or for supplementary analyses. Other microeconometric datasets used in this dissertation include student achievement data and genetic data. The main advantage of surveys in the context of this dissertation is that they allow to disclose information on attitudes, beliefs, and opinions that cannot otherwise be observed in administrative data. Surveys are not limited to elicit information about outcomes of respondents at the time of the survey, but can also ask to reveal retrospective biographical information (or follow individuals over

time in case of panel surveys). Biographical information is crucial to learn about the states and years of schooling of the respondents, which is necessary to merge data on the school curricula accordingly.

With regards to the microeconometric datasets, chapter 2 combines survey data from the National Educational Panel Study (NEPS), the German General Social Survey (ALLBUS), and the German Socio-Economic Panel (SOEP) to measure relevant outcomes of respondents from the religious sphere and beyond. Chapter 3 employs student achievement data from the National Assessment of Educational Progress (NAEP) and survey data from the General Social Survey (GSS) and the American Community Survey (ACS) to elicit knowledge, attitudes and occupational choices, respectively. Chapter 4 uses, again, student achievement data from the NAEP (from more subjects as compared to chapter 3), and combines them with teacher survey data from the NAEP and student achievement data from the Stanford Education Data Archive (SEDA) for supplementary analyses. Chapter 5 uses data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) that combine various data types including survey data, administrative data and genetic data.

1.5 Chapter Overview

This dissertation presents four empirical essays investigating various aspects of how school curricula affect students in the classroom and beyond. Each chapter contains one essay, is self-contained, and can be read independently. This section provides a concise summary of each chapter, respectively followed by a short conclusion about the policy implications of the main findings.

Chapter 2 investigates whether compulsory religious education in schools affects students' religiosity and further outcomes as adults. This chapter is joint work with Ludger Woessmann and Larissa Zierow. While the existing literature shows that religious attitudes have far-reaching consequences for individuals and societies alike, rigorous research on the determinants of such attitudes faces a challenging task as they are often deeply rooted in humans' personality and socialization. In this chapter, we ask the question whether religious attitudes can be taught in school. To address it, we exploit the staggered termination of compulsory religious education across German states in models with state and cohort fixed effects. Using three different datasets, we find that abolishing compulsory religious education significantly reduced religiosity of affected students in adulthood. It also reduced the religious actions of personal prayer, church-going, and church membership. Beyond religious attitudes, the reform led to more equalized gender roles, fewer marriages and children, and higher labor-market participation and earnings. The reform did not affect ethical and political values or non-religious school outcomes.

We conclude from our results that the content of school curricula can serve as an effective policy tool to shape important lifetime outcomes of students. The findings of our paper also

provide empirical evidence on some arguments put forward in favor or against religious education in the political debate. For example, the decrease in religious affiliation caused by the reform indicates that the historically critical stance of the churches regarding secular education was rational in the sense that forfeiting the opportunity to instill religious attitudes in public schools did undermine churches' follower base in the long run. At the same time, our finding that ethical outcomes were not affected by the reform speak against concerns in the policy debate at the time that abolishing compulsory religious education may deteriorate students' ethical orientation.

Chapter 3 provides evidence that the content of science education affects related knowledge, attitudes and choices of exposed students. The starting point for this chapter is the observation that anti-scientific attitudes can impose substantial costs on societies. While anti-scientific attitudes are generally shaped by a multitude of factors, I ask whether schools can be an important agent in mitigating the propagation of such attitudes. Specifically, this analysis exploits staggered reforms reducing or expanding the coverage of evolution theory in US state science education standards. I compare adjacent cohorts in models with state and cohort fixed effects and conduct fine-grained placebo tests to rule out scientific, religious and political confounders. There are three main results. First, expanded evolution coverage increases students' knowledge about evolution. Second, the reforms translate into evolution approval in adulthood, but do not crowd out religiosity or affect political attitudes. Third, the reforms affect high-stakes life decisions, namely the probability to work in life sciences.

In conclusion, this chapter shows that the content of education standards lastingly shapes students. This is true even for a topic like evolution that is highly charged in political and societal debates. Despite its fundamental relevance for and overwhelming acceptance in science, people in the US have strong partisan views on it. Still, what schools teach has long-run effects on individuals' fundamental views and translates into high-stakes choices. The specific findings of this chapter are of particular importance to policymakers who aim to promote scientific knowledge and attitudes in the population and increase the share of workers in STEM fields.

Chapter 4 demonstrates that reforming education standards can have unintended consequences too. This chapter is joint work with M. Danish Shakeel. From 2010 onward, most US states have aligned their education standards by adopting the Common Core State Standards (CCSS) for math and English Language Arts. Existing literature suggests that the adoption of the CCSS had zero to small positive effects on student achievement in these subjects. The CCSS did not target other subjects such as science and social studies. In this chapter, we estimate spillovers of the CCSS on student achievement in non-targeted subjects in models with state and year fixed effects. Using student achievement data from the NAEP, we show that the CCSS had a negative effect on student achievement in non-targeted subjects. This negative effect is largest for underprivileged students, exacerbating racial and socioeconomic student achievement gaps. Using teacher surveys, we show that the CCSS caused a reduction in instructional focus on non-targeted subjects.

We conclude from the findings of this chapter that the CCSS might have been more beneficial if it had been adopted for all school subjects. Such a policy might have prevented the negative spillover of the CCSS on non-targeted subjects, arguably by avoiding the perception that these subjects are less relevant and receive less instructional attention. At the same time, such a policy might also have reduced any positive effects on student achievement in the targeted subjects. In either case, adopting education standards that cover all subjects requires political challenges to be overcome as exemplified by US controversies around the history curriculum regarding the coverage of slavery or the science curriculum regarding the treatment of evolution theory and climate change.

Chapter 5 studies the interplay of genetic endowments and schooling environments in the production of educational outcomes. This chapter is joint work with Paul Hufe and Marc Stoeckli. Genetic endowments are fixed at conception and matter for the educational attainment of individuals. Our research question asks whether investments in schooling environments cushion or magnify the outcomes of this genetic lottery. Recent advances in molecular biology have opened up the black box of the human genome. Specifically, large-sample sequencing of the human genome allows to construct measures of the genetic predisposition for an outcome in question. Hence, gene-environment interactions can be analyzed more directly, going beyond indirect inference derived from twin and adoption comparisons. Using data from a representative sample of US adolescents, we analyze the interdependent associations of a measure for the genetic predisposition for educational success, teacher quality and teacher quantity with educational attainment. Our results suggest that higher-quality teachers act as substitutes for genetic endowments: a 1 SD increase in teacher quality reduces the positive association of educational attainment with the relevant polygenic score by 19 percent. In particular, with high-quality teachers genetically disadvantaged students experience a higher probability of completing college. These increases are underpinned by gains in health, language ability, patience, and risk aversion.

Teachers play a vital role for the transmission of school curricula to students. Prior research has shown that their effectiveness is a crucial input factor in the education production function. We show that teacher quality also mediates the associations of genetic endowments of students with educational attainments, leveling the playing field for all students regardless of their draw in the genetic lottery.

Taken together, the different chapters of this dissertation show that school curricula affect educational trajectories and labor market outcomes. Reforms of school curricula can be an effective tool for policy makers to change the knowledge and skills students develop in the classroom. Beyond classroom changes, school curricula can also affect beliefs and attitudes of exposed students in adulthood. Beyond attitudinal outcomes, they also shape important life decisions such as occupational choice, or the number of children, with lasting effects on key economic outcomes such as wages. In sum, school curricula exert a life-time influence on students.

2 Can Schools Change Religious Attitudes? Evidence from German State Reforms of Compulsory Religious Education*

2.1 Introduction

Religious attitudes are an important component of people's personalities and values. In the World Values Survey, 82 percent of participants belong to a religious denomination, 71 percent say that religion is important in their life, and 57 percent pray several times a week.[1] People's religiosity has important repercussions for their personal preferences, interpersonal interactions, and economic prosperity (e.g., Iannaccone (1998), Iyer (2016), and McCleary and Barro (2019)). Rigorous research on the emergence and determinants of religious attitudes, though, faces a challenging task as they are often deeply rooted in humans' personality and socialization. But can religious attitudes be taught in school? As public school curricula intervene in individuals' life course, this question addresses a core aspect of the interplay of churches and the state. In this paper, we study whether being exposed to compulsory religious education in school affects religiosity in adulthood. As churches tend to convey specific family and worldly norms, we also study effects beyond the religious sphere on family and labor-market outcomes.

We exploit the unique German setting where staggered reforms abolished compulsory religious education across states since the 1970s. The 1949 Constitution of West Germany had formally enshrined religious education as the only subject that is institutionalized as a regular subject in public schools, so that religious education was a compulsory subject in state curricula. Religious education was very intense: High-school graduates were exposed to roughly 1,000 hours of religious education over their school career – more than four times the hours of physics classes, for example (Havers, 1972). In reforms enacted at different points in time between 1972 and 2004, the different states replaced the obligation to attend religious education with the option to choose between denominational religious education and "ethics" as a non-denominational subject. A particularly interesting feature of the reforms is that the counterfactual to compulsory religious instruction is not to have *no* value-oriented instruction, but rather *non-denominational* value-oriented instruction. As a consequence, the reforms

* This chapter is joint work with Ludger Woessmann and Larissa Zierow. It is based on the paper 'Can Schools Change Religious Attitudes? Evidence from German State Reforms of Compulsory Religious Education', *CESifo Working Paper No. 9504*, 2022.

[1] Figures denote the average across the 60 countries participating in the World Values Survey in 2010-2014 (Inglehart et al., 2014). In Germany, these shares are 69, 37, and 33 percent, respectively.

allow us to identify the impact of the religious part of instruction, holding the overall exposure to value-oriented instruction constant.

Making use of the staggered adoption of the reform, our empirical model uses the variation in the abolishment of compulsory religious education across states and over time to study reform effects on outcomes in adulthood in two-way fixed effects models. Accounting for fixed effects for each state and birth year, the series of reforms provides plausibly exogenous variation in individuals' exposure to compulsory religious education that can be exploited in a difference-in-differences setting with varying timing of treatment. Effects are identified from differences in adult outcomes between cohorts within the same state that were and were not subject to compulsory religious education, relative to the differences between the same cohorts in other states that did not have reform events at the same time.

We use three datasets, each of which allows us to link religious (as well as family and labor-market) outcomes of adults to their state and time of schooling in childhood. Our merged dataset combines up to 58,000 observations of adults who entered primary school between 1950 and 2004 from the National Educational Panel Study (NEPS), the German General Social Survey (ALLBUS), and the German Socio-Economic Panel (SOEP).

Our results indicate that schools can indeed affect religious outcomes later in life. We find that the abolishment of compulsory religious education significantly decreased self-reported religiosity of affected students in adulthood. Conditional on state and birth-year fixed effects as well as individual-level control variables, religiosity of students who were not subject to compulsory religious education is 7 percent of a standard deviation lower on average compared to students who were subject to compulsory religious education. Event-study graphs show that reforming states do not have significantly different trends in religiosity in the years prior to reform compared to non-reforming states.

We find similar reductions in three measures capturing specific religious actions: the personal act of prayer, the public act of going to church, and the formal (and costly) act of church membership. Estimation of time-varying treatment effects indicates that effects on religiosity and personal prayer phase in gradually over time, whereas the effect on church membership is closer to a one-time shift. In a subsample that allows to merge regional information, effects are mostly restricted to predominantly Catholic (rather than Protestant) counties.

Beyond the religious sphere, the reforms also affected family and labor-market outcomes. First, a number of attitudinal measures indicate that the reforms led to more equitable and less conservative attitudes towards gender roles and family norms. Second, the reforms affected actual family outcomes by reducing the incidence of marriage and the number of children. Third, the reforms led to increases in labor-market participation, employment, working hours, and earnings. By contrast, there is no evidence of effects on ethical-value outcomes such as reciprocity, trust, volunteering, and life satisfaction, nor on political-value outcomes such as political interest and leaning, voting, and satisfaction with democracy. Consistent with

the counterfactual of alternative value-oriented instruction, the reform-induced decline in religiosity thus did not come at the detriment of reduced ethical values in general.

Several specification and robustness tests support our baseline result. The reforms are not related to placebo outcomes such as years of schooling, type of school degree, or age of first employment, indicating that the identifying variation is unlikely to capture alternative sources such as other contemporaneous educational reforms. Relatedly, results do not change when conditioning on a range of other educational reforms. Results are robust when restricting the sample to individuals who attend school in counties neighboring each other across state borders and including county-pair fixed effects, so that the identifying variation stems from close geographic areas. Results are also confirmed in a series of additional robustness tests and diagnostic tools of the two-way fixed effect estimator (Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021).

Our study contributes to three strands of literature. First, studies in the economics of religion have shown the importance of religion and religiosity for economic development and personal outcomes (see Barro and McCleary (2003) and McCleary and Barro (2006a, 2019) for a cross-country setting, Becker et al. (2021) for a historical context, and Becker and Woessmann (2009, 2018) for the German context). Recent analyses of the determinants of religiosity and the demand for religious services investigate, among others, effects of secular competition (Gruber and Hungerman, 2008), economic deprivation (Becker and Woessmann, 2013), printing technology (Rubin, 2014), the performance of pastors (Engelberg et al., 2016), coping with natural disasters (Bentzen, 2019), and an adult religious-value intervention (Bryan et al., 2021). Several papers study the interrelationship between education systems and religion in different contexts (Brown and Taylor, 2007; Glaeser and Sacerdote, 2008; Chaudhary and Rubin, 2011; Franck and Iannaccone, 2014; Hungerman, 2014; Meyersson, 2014; Becker et al., 2017). To the extent that they analyze effects of education on religion, these papers focus on effects of the level of education in general. Here, we focus on a different aspect – the effect of religious education in the school curriculum – as a more direct means by which schools may affect religiosity.

Second, the political economy of state schooling studies why states take over control of school curricula, modeling aspects such as totalitarian indoctrination (Lott, 1999), social cohesion (Gradstein and Justman, 2002), and socialization (Pritchett and Viarengo, 2015).[2] Historically, most Western school systems have their roots with the churches, which then exerted fierce resistance to the emerging state-sponsored non-denominational education systems during the 19th century (Ramirez and Boli, 1987; West and Woessmann, 2010).[3] Our results suggest that this resistance was rational in the sense that forfeiting the opportunity to instill religious attitudes in public schools did undermine churches' follower base in the long run.

[2] Focusing on the relationship of church and state beyond education, Barro and McCleary (2005) study determinants of state religions and McCleary and Barro (2006b) investigate their effects on religiosity.

[3] Bazzi et al. (2020) show that a backlash of Islamic schools against mass secular education increased religiosity in Indonesia in the 1970s.

Third, a broad literature in the economics of education studies the impact of different school reforms (e.g., Hanushek (1986) and Woessmann (2016)). While this literature has traditionally looked at students' academic achievement and later labor-market success, more recent contributions also focus on non-academic outcomes such as personality traits (e.g., Almlund et al. (2011)), soft skills (e.g., Koch et al. (2015)), or political attitudes (Cantoni et al., 2017). We contribute to this literature by studying how school curricula reforms can affect outcomes beyond traditional achievement measures, namely religious attitudes in the long run.

In the following, section 2.2 provides institutional background on the studied reforms. Section 2.3 describes the empirical model and section 2.4 the data. Sections 2.5 and 2.6 present our results on reform effects on religious outcomes and on family and labor-market outcomes, respectively. Section 2.7 reports specification and robustness tests, and section 2.8 concludes.

2.2 Institutional Background: Reforms Abolishing Compulsory Religious Education in Germany

With the staggered abolishment of compulsory religious education across states and over time, Germany provides a unique setting to study the effects of compulsory religious education.[4]

2.2.1 Historical Background

There are a couple of historical milestones that led to the profound role of religious education in the German school system. The Prussian School Supervision Act of 1872 was at the center of the *Kulturkampf* ("culture battle") between the Prussian Empire under Bismarck and the Catholic Church during the 1870s. This legislation abolished the churches' control of the Prussian primary school system, putting the state in charge of school organization and curricula with the aim to provide a value-neutral education. However, religious education remained a regular school subject. During the Weimar Republic (1918-1933), there was some debate about whether religious education should be offered in schools at all, but in the end the supporters of religious education prevailed.

In Nazi Germany, the role of religious education was formally strengthened by the *Reichskonkordat* (Concordat between the Holy See and the German Reich) closed between Hitler and the Pope. It assigned Catholic religious education the role of a regular school subject. In reality, however, the Nazi regime did not adhere to these rules. A prominent example is the so-called *Kreuzkampf* ("cross battle") in the region of Oldenburg Münsterland in 1936, where the regional minister for education and church gave the order to take away all crosses, pictures, and other religious symbols from schools (*Kreuzerlass*). After protests by civil society

[4] By contrast, it is hard to imagine exogenous variation in religious education in countries where it is barred from public schools (e.g., the United States with its strict separation of church and state that forbids religious education in public schools) or offered as an elective subject (e.g., Italy or the Netherlands).

that were famously supported by Bishop Clemens August Graf von Galen, the order was partly taken back, and crosses were again allowed to be placed in schools in this region. Referring back to Bishop Wilhelm Emmanuel von Ketteler during the *Kulturkampf*, Bishop von Galen strongly emphasized the crucial role of the church's (rather than the state's) grip of schools for the children's socialization and thus for church followership in the long run.

2.2.2 Post-War Situation

Against the backdrop of the Nazi takeover of schools and in close agreement with the Allied forces, the Constitution (*Grundgesetz*) of the Federal Republic of Germany, enacted in 1949, establishes in Article 7 that religious education is a regular subject in public schools.[5] This makes it compulsory that public schools provide religious education, which is explicitly to be taught in accordance with the principles of the respective religious community. Before reforms that started in the 1970s, enrollment in religious education classes was the default for all students from first to final grade. Parents (and adolescents aged at least 14)[6] could formally request non-participation if the child was not baptized, but this was a rare exception (Havers, 1972).

Children are taught by confession (Catholic or Protestant). Based on contracts between the states – who are responsible for education policy – and the churches, the content is not restricted to "religious studies" but is based on dogmatic elements bound to the respective denomination and its doctrinal theology (Lott, 2005). Religious-education teachers are paid by the states and work as state employees but must be chosen and certified by the respective church (receiving the Catholic *Missio canonica* or the Protestant *Vocatio*). The importance given to the subject in Germany's school curricula is illustrated by the fact that during their school careers, high-school graduates were exposed to 1,000 hours of religious education – compared, e.g., to 240 hours of physics education (Havers (1972) based on the Baden-Wuerttemberg curriculum).

2.2.3 The Reforms

From the 1970s onwards, eight of the eleven West German states terminated the compulsory nature of religious education (Helbig and Nikolai, 2015). Parents could now choose between religious education and a newly introduced subject, usually called "ethics",[7] which provides an alternative form of value-oriented instruction that was non-denominational. As indicated in Table 2.1, Bavaria was the first state to enact the reform in 1972 and Hamburg and North Rhine-Westphalia were the last in 2004 (see also map in Appendix Figure A2.1).

[5] Article 141 states that this clause does not apply to states that had had a different state law on the issue in place on January 1, 1949, which effectively granted an exemption to the two city states of Berlin and Bremen.

[6] In Bavaria and the Saarland, students had to get parental permission until age 18.

[7] Depending on the state, the alternative subject is called "ethics", "philosophy", "values and norms", or "humanistic life skills".

Two reasons are generally put forward for the reform introduction, one on the initiative of the churches and the other of the schools (Havers, 1972; Lott, 2005). First, in 1968 the student movement at German universities started to challenge tradition and conservatism of the parental generation. When an increasing number of high-school students in urban areas decided to opt out of religious education to enjoy free time, the churches reacted by pushing for a compulsory alternative subject that students are obliged to attend instead, to make opt-out less attractive.[8] Consistent with the initiation by the churches, Bavaria – which in many dimensions is generally viewed as the most conservative among the West German states – was the first to enact the reform. Second, schools also welcomed the reform, as rising opt-out meant that they were increasingly faced with organizational challenges to comply with their supervisory duty for students during school hours.

Interestingly, the rollout of the reform across states was orthogonal to the political leaning of and changes in the state government. As is obvious from column 4 of Table 2.1, four reforms were implemented by a right-of-center Christian Democrat (CDU/CSU) government and four by a left-of-center Social Democrat (SPD) government. The time pattern is literally alternating between the two camps. Furthermore, for each single reform, the party that was in power in the legislative period of the reform had already been in power in the prior legislative period, implying that no reform was implemented after a change in government (column 5). Similarly, the reform rollout was not driven by the size of a state, as the two largest states (Bavaria and North Rhine-Westphalia) were the first and last to implement the reform, respectively. These patterns make it unlikely that the reforms were due to political trends or shocks.[9]

There are three main consequences of the reform that might give rise to overall long-term reform effects. First, individual students could now attend ethics instead of religious education. Unfortunately, there is no administrative data on how many students chose ethics in the years right after the reform implementation. Initially, the number was potentially small, particularly in rural areas. Reports dating back to the reform years suggest that in some places, schools could not find staff to teach ethics classes (Lott, 2005). Selective data in later years point towards a modest decline in the number of students attending religious education. Current data indicate that 73 percent of students in West German public schools attend religious education and 20 percent ethics or related substitute subjects (Kultusministerkonferenz,

[8] To ensure that results are not driven by reactive reforms to early opt-out during the student movement, in robustness tests we show that results are robust to leaving out early reforming states and to restricting the sample to rural areas (see section 2.7).

[9] The result that we do not find reform effects on political outcomes (section 2.5) also speaks against the existence of political shocks coinciding with the timing of the reforms.

2021).[10] Thus, only about one fifth of students are affected in the sense that they themselves attend non-denominational ethics rather than denominational religious education.

Second, the subject ethics acted as a newly emerged competitor to religious education, putting religious education curricula under modernizing pressure. Studying curricula before and after the reform, we find that religious education curricula tended to change after the reform. As one example, Appendix Table A2.1 provides an overview of curricula in Bavaria. The 1967 pre-reform curriculum of Catholic religious education never even mentions non-Christian religions. By contrast, the 1979 post-reform curriculum has a whole section in grade 9 designated to learning about other religions. The pre-reform curriculum puts more focus on guiding students towards Christianity, whereas the post-reform curriculum emphasizes guiding students towards responsible and informed behavior defined by Christian values.[11] As an example of a late reformer, the 1999 pre-reform syllabus in North Rhine-Westphalia focuses on religious values to guide students, whereas the 2014 post-reform syllabus emphasizes helping students develop their own values based on religion and faith. Overall, the comparison of curricula points to a decrease in the practice of prayers and literal interpretation of the bible after the compulsory nature of religious education was abolished.

Third, the reform may have changed perceived social norms since it was now officially approved that alternatives to religious education exist, indicating an apparent acceptance in society not to be religious. This could have changed religious views even for students who still attended religious education classes. To the extent that these effects are specific to the affected student cohorts rather than to the population overall, they would be captured by our empirical approach.

Any identified long-term reform effects are therefore likely to stem from a combination of declining attendance in religious education, adapting the content of religious education classes to the new competitor subject's content, and changing social norms. We therefore expect that the reform does not only affect students who chose to attend ethics classes, but also students who continued attending religious education classes. In addition, the description makes clear that several elements of the enactment of the reform were gradual

[10] The number for religious education includes all religions (including Islam and Judaism) as well as denomination-overarching religious education; 33 percent of West German students attend Catholic and 34 percent Protestant religious education. 7 percent of students attend neither religious education nor ethics, which mostly refers to primary schools in North Rhine-Westphalia, where ethics is not yet ubiquitously implemented in all schools, and to secondary schools in Schleswig-Holstein, where religious education/ethics classes of consecutive grades can be offered combinedly in one grade so that students in the other grade currently do not attend it.

[11] In the syllabus of the new subject ethics in Bavaria, religion of any kind is completely absent (except for one reference to Christian values). The focus is on enabling students to work out answers to ethical questions by themselves in open discussions based on real-life situations. After the curricular changes in religious education, ethics and religious education have a lot of common topics and focus both on conveying values; the major difference is the final justification of values taught in class (Schwoerbel, 1985).

rather than abrupt, leading to an expectation that reform effects may phase in rather than happen discontinuously.

2.3 Empirical Model

To estimate the effect of the abolishment of compulsory religious education on religiosity and other outcomes in adulthood, we make use of the different timing of reform events across German states. The staggered adoption of the reform allows us to estimate reform effects in a generalized difference-in-differences setting with varying timing of treatment. The key idea is that states without a reform in a certain year act as counterfactuals for states with a reform in that year, after accounting for time-invariant differences between states and national differences between years. Our baseline two-way fixed effects model with state and cohort fixed effects models reform effects as immediate and permanent shifts in outcomes in the reforming states and years, relative to non-reforming states and years:

$$R_{i,s,t} = 1(t_{i,s} \geq t_s^*)\beta_{Reform} + \mathbf{X_i}\beta_{Controls} + \mu_s + \lambda_t + \epsilon_{i,s,t} \tag{2.1}$$

The adulthood outcome (e.g., religiosity) $R_{i,s,t}$ of individual i who started primary school in state s and year t is a function of an indicator term $1(t_{i,s} \geq t_s^*)$ that equals one if the primary school entry year $t_{i,s}$ of individual i in state s is larger than or equal to the year of reform t_s^* in state s.[12]

Apart from state and cohort fixed effects (μ_s and λ_t, respectively), a vector of individual-level controls $\mathbf{X_i}$ and an error term $\epsilon_{i,s,t}$ complete the model. Throughout the paper, standard errors are clustered at the state level. We report p-values for two clustering methods. The first one is the standard clustering approach which accounts for potential correlation of error terms across years within states and provides conservative inference if reform timing is random (Abadie et al., 2017; Athey and Imbens, 2021). The second one is the wild cluster bootstrap approach suggested by Roodman et al. (2019) which provides asymptotic refinement by accounting for the limited number of clusters given by the West German states.[13]

[12] Coding individuals as treated only if the reform had been implemented at their primary school entry is our preferred categorization because it starts with the first cohort that could have avoided religious education completely by choosing the non-denominational alternative from the first grade onwards. The fact that students who were already beyond primary school entry in the year of reform introduction are categorized as exposed to compulsory religious education even if they received some exposure to the reformed curriculum might introduce attenuation bias in our baseline specification. In robustness analyses, we confirm results in a dosage specification where treatment is defined as the share of compulsory school years that an individual spent in the reformed system, as well as in a specification that defines treatment by entry into secondary school (see section 2.7).

[13] We use Webb weights and 9999 replications. The approach is more conservative than the Cameron et al. (2008) approach to wild cluster bootstrapping which tends to yield substantially lower p-values throughout (not shown).

The parameter of interest, β_{Reform}, depicts the intention to treat (ITT) effect that captures the overall effect of the reform, that is, the effect of being offered the choice between attending religious education or ethics. The treatment effect is identified from changes in adult outcomes across cohorts within the same state that were and were not affected by the reform, relative to the same changes in other states without reform events at the same time.

The variation in the timing of reforms across states provides us with plausibly exogenous variation in individuals' exposure to compulsory religious education. The main identifying assumption is that the exact timing of the reform is as good as random (e.g., Athey and Imbens (2021) and Borusyak et al. (2021)). This seems plausible given the idiosyncrasies of the reform processes in the German federal political system described above. For example, the reform rollout did not indicate any political trend, with implementations alternating between right-wing and left-wing governments and no reform enacted in the first legislative period after a change in government (see Table 2.1).

One way in which the identifying assumption could be violated is the existence of other school reforms that happened simultaneously. However, the timing of the religious-education reform is very peculiar, and we are not aware of other reforms with even vaguely similar patterns of timing across states. In fact, results are robust in specifications that control for a range of other education reforms (see section 2.7). An additional way to test this concern is to estimate reform effects on non-religious school outcomes such as degree completion or years of schooling. The religious-education reform did not affect any other subjects and did not substitute religious education by classes prone to enhance achievement in other curricular subjects. As we thus do not expect any effects of the religious-education curriculum on other school outcomes, such analysis can be interpreted as a placebo test that, if it failed, would indicate the possibility of simultaneous school reforms.

In a further specification that aims to compare observations that are as similar as possible in the absence of treatment, we restrict the sample to individuals living in counties that are directly at the border to a different state. In this specification, we can additionally include fixed effects for each pair of counties that are next to each other on either side of a state border, thereby further reducing geographic heterogeneity in the identifying variation.[14]

In addition, it is an attractive feature of the event-study approach that including a trend variable relative to the reform ($t_{i,s} - t_s^*$) constitutes a falsification test of the identifying assumption of randomness in reform timing (keeping the assumption of time-invariant treatment effects for now):

$$R_{i,s,t} = 1(t_{i,s} \geq t_s^*)\beta_{Reform} + (t_{i,s} - t_s^*)\beta_{Trend} + \mathbf{X_i}\beta_{Controls} + \mu_s + \lambda_t + \epsilon_{i,s,t} \qquad (2.2)$$

[14] Counties (*Landkreise and kreisfreie Städte*) in Germany are substantially smaller than in the US. There are 325 counties in West Germany with a mean population of about 200,000 inhabitants (median about 150,000).

The parameter β_{Trend} captures how the average outcomes change in reforming states relative to non-reforming states. Rejecting the null hypothesis that $\beta_{Trend} = 0$ would indicate that the timing of the reform may not be as good as random.

While specifications (1) and (2) model the reform as an immediate and permanent shock, the discussion in section 2 suggests that reform implementation may have been gradual rather than abrupt. To disentangle reform effects that happen directly at the time of the reform from those that occur gradually afterwards, we extend specification (2) by an interaction of the reform indicator ($t_{i,s} \geq t_s^*$) with the trend term ($t_{i,s} - t_s^*$):

$$R_{i,s,t} = 1(t_{i,s} \geq t_s^*)\beta_{Reform} + (t_{i,s} - t_s^*)\beta_{Trend} + 1(t_{i,s} \geq t_s^*)(t_{i,s} - t_s^*)\beta_{Reform*Trend} + \mathbf{X_i}\beta_{Controls} + \mu_s + \lambda_t + \epsilon_{i,s,t} \tag{2.3}$$

In this specification, the parameter on the interaction term, $\beta_{Reform*Trend}$, captures the average annual change in the outcome in reforming states after the reform, relative to the average annual change in the same states prior to the reform (and relative to non-reforming states). The parameters β_{Reform} and $\beta_{Reform*Trend}$ reveal whether the reform affects outcomes as immediate permanent shocks or gradually over time, respectively (Lafortune et al., 2018). The parameter β_{Trend} now captures differential pre-trends between treatment and control states.

To lift the assumption of linearity in pre- and post-trends of the parametric specifications and allow for flexible reform effects over time, we also estimate non-parametric models of the effects of a reform in year t_s^* on outcomes k years before and after the reform:

$$R_{i,s,t} = \sum_{k=-19}^{20} 1(t_{i,s} = t_s^* + k)\beta_k + \mathbf{X_i}\beta_{Controls} + \mu_s + \lambda_t + \epsilon_{i,s,t} \tag{2.4}$$

Effects, captured by the parameter vector β_k, are estimated relative to the excluded category $k = 0$. To smooth the numbers of observations in the sample across years, we group observations together to bins of five years each. We visualize the results of this non-parametric specification in an event-study graph.

The two-way fixed effects model assumes homogeneity in treatment effects (e.g., Sun and Abraham (2021)). We implement the estimator suggested by Callaway and Sant'Anna (2021) and use the diagnostic tools suggested by Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021) to show that our results are not contaminated by this assumption.[15]

[15] Furthermore, excluding covariates does not change our qualitative results, indicating that cohorts with different covariates are unlikely to react differently to the reform (see Appendix Table A2.21).

2.4 Data

Our treatment variable indicates whether a given German state has abolished compulsory religious education at a given point in time. The coding of reform events, indicated in Table 2.1, is taken from Helbig and Nikolai (2015). We define an individual as treated if the reform that replaced compulsory religious education by the choice between ethics and religious education had been enacted in the year that the individual entered primary school.

To estimate reform effects on individuals' adult outcomes, we looked for individual-level datasets that provide a broad picture of religiosity in Germany. We ended up using three different datasets that are each drawn to be representative for the German adult population (see Data Appendix A2.1 for details): the adult cohort of the National Education Panel Study (NEPS), the German General Social Survey (ALLBUS), and the German Socio-Economic Panel (SOEP). NEPS is focused on the educational sciences and provides a panel of over 12,000 adults observed between 2007 and 2016. ALLBUS is focused on the social sciences and provides repeated cross-sections of over 15,000 adults observed between 1980 and 2016. SOEP is focused on economics and the social sciences and provides a panel of over 30,000 adults observed between 1984 and 2017. To study a range of religious (and other) outcomes in adulthood and maximize statistical power, we use all three datasets and merge them together in our main analysis. Depending on the outcome under study, our combined estimation sample includes up to 58,000 observations.

All three datasets meet the basic data requirement of our evaluation approach that we need to observe individuals' state and year of primary school entry. Each dataset thus allows us to link the religiosity of individuals in adulthood to their state of schooling in childhood, even if they migrated to other states in-between.[16] Our sample consists of all individuals who entered primary school in West Germany between 1950 and 2004. We exclude individuals who entered primary school before 1950 because they did not have their entire schooling career in the Federal Republic of Germany (founded in 1949). Primary school entry by 2004 ensures that individuals have reached adulthood by 2016/17.

Our main outcome of interest is self-reported religiosity, which we interpret as a comprehensive measure describing an individual both believing in religious content and living a religious life in public, i.e., showing religious belonging (following the terminology of McCleary and Barro (2019)).[17] The three other religious outcome measures capture different ways in which individuals articulate their religiosity in specific actions: the personal act of prayer, the public act of going to church, and the formal act of church membership. The latter act is also directly economically relevant, as church membership in Germany is automatically related to paying church taxes (levied as a surcharge on income tax).

[16] If available directly, we use information on the year and state of primary school entry. If not, we use the year and state of birth and assume that individuals enter primary school six years later in the same state.

[17] The available data do not allow us to cleanly distinguish between the believing and belonging dimensions of religiosity.

As the religious outcome variables are elicited with varying numbers of answer categories in the different datasets (see Appendix Table A2.2 for details), we standardize the religious measures within each dataset before merging the three surveys together and include dataset fixed effects throughout.[18] If a measure is observed multiple times per individual in a panel dataset, we use the most recent available observation on any given variable and include survey-year fixed effects (stored separately for each question for each individual) throughout.

The three datasets also provide batteries of measures of attitudes towards gender and family roles and of actual family and labor-market outcomes, as well as of ethical-value, political-value, and educational outcomes. Control variables include gender, migration status, and mothers' and fathers' education. Table 2.2 provides descriptive statistics for the merged dataset. Roughly one third of observations are treated by the reform, i.e., they entered primary school after compulsory religious education had been abolished.

2.5 The Effect of Abolishing Compulsory Religious Education on Religiosity

This section reports our baseline results on effects of the studied reform on religious outcomes. Section 2.6 turns to effects on non-religious outcomes, and section 2.7 provides results of specification and robustness tests.

Our results show that the abolishment of compulsory religious education decreased the religiosity of affected students in adulthood. The event-study graph of Figure 2.1 indicates that individuals who entered school after the reform report significantly lower levels of religiosity.[19] Visual inspection suggests that reform effects appear to phase in gradually over time. An omnibus hypothesis test that the post-event effects are jointly zero is rejected at the 1 percent level. By contrast, the test does not reject that the pre-event effects are jointly zero, indicating that reforming states had not been on different trends from non-reforming states prior to the reform.

The parametric estimation in the first column of Table 2.3 indicates that reform exposure while being in school decreases religiosity in adulthood by 7 percent of a standard deviation on average. For a straightforward indication of the magnitude of this effect, we can express religiosity as a dummy variable. The reform reduces the likelihood that a person is (rather or very) religious by 2.9 percentage points (independent of whether estimated by linear probability or probit model; see Appendix Table A2.4), compared to an average incidence of 52.4 percent in our dataset. The incidence of being very religious is reduced by 2.2 percentage points (compared to an average incidence of 10.9 percent).

[18] To document that results are not driven by the standardized merging, robustness checks also show results for each of the three datasets separately (see section 2.7).

[19] Appendix Table A2.3 provides the non-parametric regression results underlying this figure.

The reform also led to significant reductions in the three measures of specific religious actions (columns 2-4 of Table 2.3). The standardized effects are of a similar magnitude to overall religiosity. The reform reduces the personal act of prayer by 5 percent of a standard deviation (marginally significant), the public act of going to church by 7 percent, and the formal act of church membership by 8 percent.[20]

To test whether reforming states are on a general time trend that is different from non-reforming states, the odd columns of Table 2.4 add a linear trend relative to the respective reform event to the model. There is no significant differential trend for religiosity or any of the religious-action outcomes, in line with the assumption that the timing of reform events is as good as random.

The even columns of Table 2.4 report results of the rather demanding specification with time-varying treatment effects that allows for both a shift term of the reform, a relative trend, and an interaction between the two. Confirming the graphical depiction, results indicate that the reform effect on religiosity phases in gradually over time: religiosity decreases by 0.013 standard deviations on average per year in reforming states after the reform, relative to the average change in the same state prior to the reform. A similar gradual treatment effect emerges for personal prayer. By contrast, the effect on affiliation with a religious community is mostly captured by a one-time shift. This may be related to the fact that church membership in Germany implies the requirement to pay church taxes: Individuals who were exposed to the reform even in the early years after a state's implementation do react by leaving their church as adults to avoid paying church taxes, whereas their subjective religiosity and prayer are not yet as strongly impacted. For church-going, the separate estimates in this specification are too imprecise to distinguish between a one-time shift and gradual phasing-in.

Treatment effects on religiosity are very similar for women and men (Panel A of Table 2.5). The same is true for church affiliation. By contrast, treatment effects on prayer only materialize for females but not males, whereas treatment effects on church-going are larger for males. Results in Panel B show no strongly differential pattern for individuals who went to schools in rural and urban areas (available for a limited number of observations in RemoteNEPS). The effect is somewhat larger (although less precisely estimated) in urban areas for religiosity, larger in rural areas for prayer, and similar for affiliation. When distinguishing individuals' school county by the majority confession (Panel C), results are driven by Catholic areas. In another subset of observations and outcomes (available in ALLBUS and SOEP) where we can link individuals to the denomination of their parents (Panel D), the effect on church-going also appears to be restricted to individuals with all-Catholic parents. By contrast, while estimates are somewhat imprecise, the effect on religious affiliation is in fact larger for individuals whose both parents were Protestant.

[20] Appendix Figures A2.2-A2.4 show the respective event-study graphs.

In contrast to the effects on religiosity and religious actions, we do not find evidence that the reform affected various value outcomes. In particular, there are no significant treatment effects on a series of measures of ethical-value outcomes including reciprocity, trust, risk preference, volunteering, and life satisfaction (Panel A of Table 2.6). The absence of treatment effects on these ethical outcomes is consistent with the fact that the counterfactual to compulsory religious education in our setting is not no value-oriented classes, but rather a choice between two types of value-oriented classes that are either denominational or not. Apparently, attending the non-denominational subject ethics does not lead to lower levels of the different ethical-value outcomes compared to the subject religious education. Similarly, there is no evidence of effects on political-value outcomes such as political interest, satisfaction with democracy, or left-right voting patterns (Panel B).

2.6 Effects on Family and Labor-Market Outcomes

Historically, the churches strongly promoted traditional religious family role models, advocating gender-specific roles in families and marriage before cohabitation. Therefore, we also study effects of the termination of compulsory religious education beyond the religious sphere on people's attitudes towards gender and family roles and on subsequent family and labor-market outcomes.

Results show that the reform led to a decrease in conservative gender and family attitudes. Abolishing compulsory religious education reduced the likelihood to think that men are better suited for certain professions than women by 8 percent of a standard deviation (column 1 of Panel A of Table 2.7). Results on views on equal gender duties in housework are shy of statistical significance (column 2), but the reform also significantly decreased the likelihood to think that women cannot use technical devices as well as men (column 3). Similarly, the reform reduced the view that people should get married if they permanently live with a partner (column 4).

The reform also affected actual family outcomes. The treatment reduced the probability to be married by 1.5 percentage points (column 1 of Panel B of Table 2.7), compared to an average marriage rate of 60 percent. The reform also decreased the number of children by 0.09 children per respondent (column 2), compared to an average of 1.4 children.

The reform may additionally have affected economic behavior and outcomes. According to Christian values, the decrease in religiosity may have promoted materialistic orientation.[21] The reduction in time used for various religious actions may have induced a substitution effect towards economic activities (Barro and McCleary, 2003; Gruber and Hungerman, 2008). The reduced time required to raise (fewer) children may also have changed decisions about family

[21] For example, the bible quotes Jesus as saying, "It is easier for a camel to go through the eye of a needle than for someone who is rich to enter the kingdom of God." (Mark 10:24-27, Luke 18:24-27)

and career planning, and the changed gender roles may have opened up better labor-market opportunities for women. Finally, leaving the church means a reduction in the tax rate on labor income in Germany, increasing incentives to work.

Results show that the reform indeed had positive effects on labor-market outcomes. The probability to participate in the labor market increases by 1.5 percentage points (column 3 of Panel B of Table 2.7), compared to a mean of 82 percent, and the probability to be employed by 2.3 percentage points (column 4; mean 78 percent). Among those employed, working hours rise by 0.6 hours per week (column 5), compared to a mean of 35.6 hours. Earnings increase by 5.3 percent (column 6). Overall, the results suggest that the reform impacted people's lives well beyond the religious sphere.[22]

2.7 Robustness

This section reports tests of challenges to our identification strategy, of the robustness of our results, and of properties of the two-way fixed effects estimator.

2.7.1 Placebo test: Effects on Non-Religious School Outcomes

For our identification strategy to hold, the abolishment of religious education should not be accompanied by other educational reforms or other state-specific events with the same timing structure. As meaningful other school reforms should leave traces in general educational outcomes, one way to test this is to estimate treatment effects on non-religious educational outcomes. Results show that the reform is not significantly related to the non-religious educational outcomes in our datasets, namely years of schooling, the type of school degree, or the age of first employment (Table 2.8). As the studied reform did not lead to a change in schooling hours and or in the structure or content of the non-religious subjects, we interpret this as a placebo test that is in line with our identifying assumption. This interpretation is also consistent with the non-existence of effects on ethical-value and political-value outcomes (see section 2.5 above).

2.7.2 Border Specification with County-Pair Fixed Effects

To reduce the possible incidence of unobserved differences, we can restrict the analysis to individuals from geographically close and thus arguably highly similar counties. In a subset of individuals observed in the NEPS data, we observe individuals' county of schooling. This allows us to restrict the sample to pairs of counties separated by a state border (see Appendix Figure A2.5). Additionally, in this specification we can include county-pair fixed effects for each pair of neighboring counties that is divided by a state border (Dube et al., 2010; Bentzen and Sperling, 2020). The identifying variation is thus restricted to a comparison of pairs of

[22] We do not find that any of the family and labor-market effects differ significantly by gender (not shown).

counties on either side of the respective state border. In this smaller sample, the treatment effect on religiosity remains highly significant and increases in size to 0.16 standard deviations (Table 2.9). The same is true for prayer, whereas the effect on affiliation does not hold in this specification.

2.7.3 Additional Robustness Analyses

A series of additional tests confirm the robustness of our findings to variations in control variables, treatment specifications, and estimation samples. To ensure that the estimated reform effects do not pick up effects of other education reforms, we include controls for a range of other reforms. These include reforms of the length of compulsory schooling (e.g., Pischke and von Wachter (2008) and Cygan-Rehm (2021)), of the duration of the highest-track school ("G8/G9 reform", e.g., Andrietti and Su (2019) and Marcus and Zambre (2019)), and of whether philosophy, sexual education, and political education, respectively, are taught in school (see Helbig and Nikolai (2015)). Results are robust to controlling for these other education reforms (Appendix Tables A2.5-A2.8).

A couple of robustness checks relate to the coding of treatment. First, we replace the dummy variable indicating reform exposure by a dosage variable measuring the share of school years out of the total compulsory school years in which individuals were exposed to the reform. Results are robust and point estimates become larger for each of the religious outcomes (Appendix Table A2.9), as expected if the conservative baseline indicator coding suffers from attenuation bias. Second, we alternatively define a student to be exposed if the reform was in place at the time of entry into secondary (rather than primary) school, with very similar results (Appendix Table A2.10).

A potential concern in our setting is that the effects might be related to the student movement in the late 1960s and early 1970s. To test this, we exclude all early reforming states from the sample and keep only those states which reformed since the 1980s. Results are largely unaffected in this smaller sample (Appendix Table A2.11).

While our baseline analysis merges the NEPS, ALLBUS, and SOEP datasets to maximize statistical power, we also estimate the models separately for the three datasets to ensure that results are not driven by any specific dataset or by the merging. Results indicate that the effects tend to materialize in each of the separate datasets, although obviously at lower levels of statistical precision (Appendix Tables A2.12-A2.14).

2.7.4 Tests of the Two-Way Fixed Effects Estimator

Our setting generalizes the classic two-group/two-period difference-in-differences setting in that there are eleven states among whom eight change their treatment status in different years over an extended time horizon. To ensure that our estimates are not driven by two-by-two reform estimates with negative weights, we implement the estimator suggested by Callaway

and Sant'Anna (2021) that is immune to bias from negative weighting. The procedure uses only not-yet treated units and never-treated units as controls. Already-treated units, which could potentially cause negative weighting, are omitted from the analysis. Reassuringly, the aggregated estimates of the average treatment effect on the treated (ATT) for the four religious outcomes are very similar to our baseline two-way fixed effects estimates (see Appendix Table A2.15). In fact, the ATT estimates are larger (in absolute terms) than the corresponding baseline estimates, although sometimes at lower levels of statistical significance. Appendix B reports additional diagnostic tests suggested by Chaisemartin and D'Haultfœuille (2020) and by Goodman-Bacon (2021) which further corroborate our baseline results and indicate that our findings are not driven by a setting that would give rise to negative weights.

2.8 Conclusion

Our study investigates whether compulsory religious education affects people's religiosity in the long run. We argue that the different timing of reforms that abandoned compulsory religious education across German states provides plausibly exogenous variation in individuals' exposure to compulsory religious education. Students could now choose to attend non-denominational ethics classes rather than religious education, which likely also changed overall social norms towards religion and, by competitive pressures, the content of religious classes. We find that, conditional on state and birth-year fixed effects, the termination of compulsory religious education led to a significant reduction in the religiosity of affected students in adulthood. The reform reduced the share of people reporting to be religious by about 3 percentage points (compared to an average incidence of 52 percent) and of those reporting to be very religious by 2 percentage points (average 11 percent). Similar standardized reductions are found in three measures of religious actions – prayer, church-going, and religious affiliation.

We do not find that the reform affected ethical values and behavior such as reciprocity, trust, volunteering, and life satisfaction, nor political values and behavior such as interest in politics, satisfaction with democracy, or voting. It appears that the counterfactual of attending non-denominational ethics classes was equivalent to attending religious-education classes in terms of these outcomes, speaking against concerns in the policy debate at the time that abolishing compulsory religious education may deteriorate students' ethical orientation.

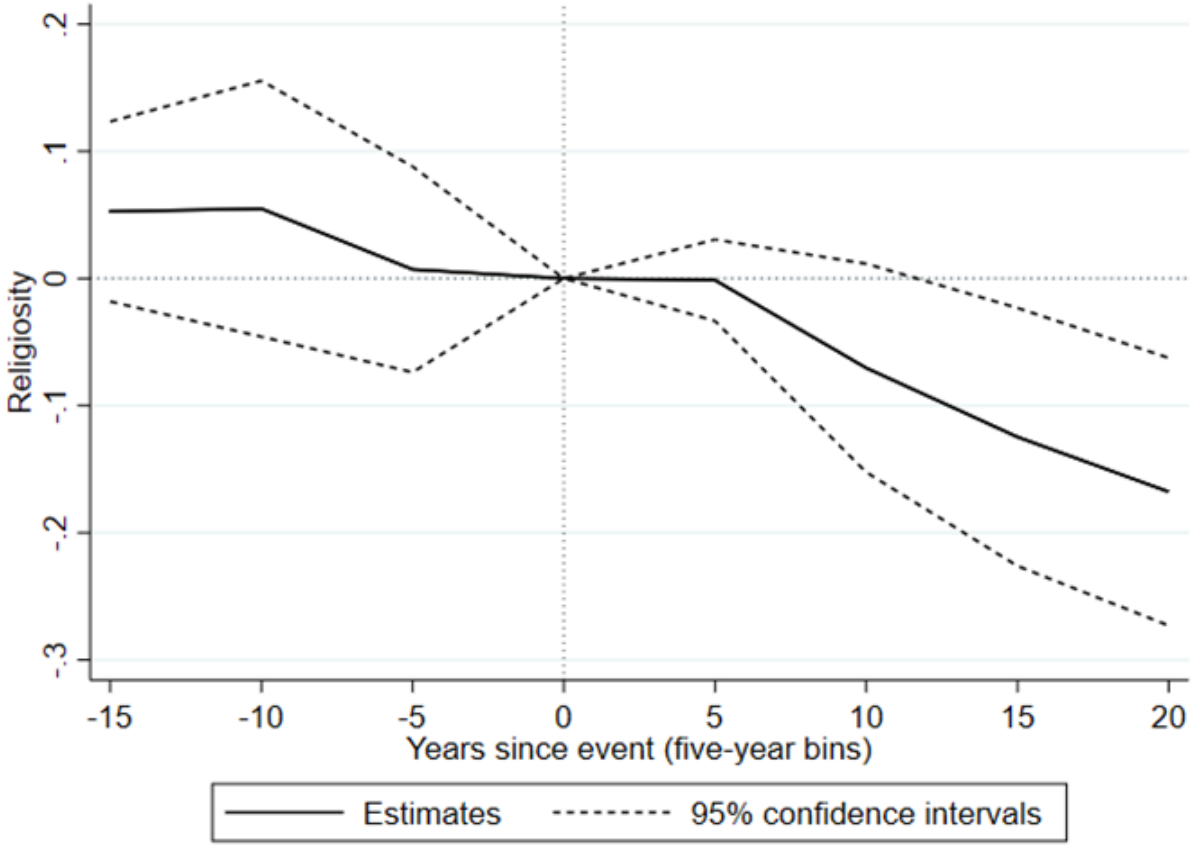
Beyond the religious sphere, the reform also affected family and economic outcomes. Affected students express less conservative gender and family norms later in life. This finding provides insights for the literature on gender norms which shows that these norms are important determinants for lifetime outcomes (e.g., Kleven et al. (2019) and Jayachandran (2021)). Yet, it is not well understood where these norms come from. Our results provide evidence that changes in school curricula can impact gender norms, implying that they are malleable in public settings outside the family. The abolishment of compulsory religious education also affected actual family outcomes – lower incidence of marriage and number of children – as

well as labor-market outcomes – higher employment and earnings. Thus, the reform also had economically relevant repercussions.

Overall, our results indicate that religious indoctrination in school can indeed exert a life-time influence on students.

Figures and Tables

Figure 2.1: The effect of abolishing compulsory religious education on religiosity: Non-parametric event-study estimates



Notes: Coefficients from non-parametric event-study regressions and their 95 percent confidence intervals. Dependent variable: religiosity (standardized, based on 4-point-scale NEPS question "How religious are you?" and 10-point-scale ALLBUS question "Would you say that you are rather religious or rather not?"). Numbers on horizontal axis refer to final year of respective five-year bins; i.e., 0 = last five years prior to treatment (excluded category), 5 = first five years of treatment. Inference: Standard clustering at state level. The p-values of omnibus hypothesis tests of zero pre- and post-event effects are 0.343 and 0.008, respectively. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table 2.1: The rollout of abolishing compulsory religious education: Timing of treatment and governing parties

			Governing parties in legislation period	
Reform year	State	Time in treatment	before the reform	of the reform
(1)	(2)	(3)	(4)	(5)
Before 1949	Berlin	1		
Before 1949	Bremen	1		
1972	Bavaria	0.60	CDU (1966-1970)	CDU (1970-1974)
1974	Lower Saxony	0.56	SPD (1970-1974)	SPD, FDP (1974-1976)
1977	Rhineland-Palatinate	0.51	CDU (1971-1975)	CDU (1975-1979)
1977	Hesse	0.51	SPD, FDP (1970-1974)	SPD, FDP (1974-1978)
1983	Baden-Württemberg	0.40	CDU (1976-1980)	CDU (1980-1984)
1992	Schleswig-Holstein	0.24	SPD (1988-1992)	SPD (1992-1996)
2004	Hamburg	0.02	CDU, PRO, FDP (2001-2004)	CDU (2004-2008)
2004	North Rhine-Westphalia	0.02	SPD, Grüne (1995-2000)	SPD, Grüne (2000-2005)
No reform	Saarland	0		

Notes: The table lists the dates of reforms abolishing compulsory religious education for the respective states (from Helbig and Nikolai (2015)), the share of years each state spends treated in the estimation sample from 1950-2004, and the governing parties before and during the reform.

Table 2.2: Descriptive statistics

	Mean	Std. Dev.	Min.	Max.	Obs.
Reform (treatment indicator)	0.32	0.47	0.00	1.00	58,703
Religious outcomes					
Religiosity	0.00	1.00	-1.69	1.77	15,688
Prayer	0.00	1.00	-1.26	2.44	13,276
Church-going	0.00	1.00	-1.16	3.07	42,776
Affiliation	0.00	1.00	-2.22	0.57	45,925
Ethical-value outcomes					
Reciprocity	0.00	1.00	-5.11	0.97	21,150
Trust	0.00	1.00	-2.71	2.01	37,070
Risk-taking	0.00	1.00	-3.00	2.64	35,556
Volunteering	0.43	0.49	0.00	1.00	37,971
Life satisfaction	0.00	1.00	-4.85	1.56	48,177
Political-value outcomes					
Interest in politics	0.00	1.00	-2.47	2.00	52,970
Politics too complicated	0.00	1.00	-1.95	2.25	9,160
Satisfaction with democracy	0.00	1.00	-2.86	1.90	14,519
Political spectrum: right	0.00	1.00	-3.02	3.37	40,161
Vote in election	0.87	0.34	0.00	1.00	32,133
Vote left	0.57	0.49	0.00	1.00	27,088
Vote extreme	0.07	0.25	0.00	1.00	27,100
Attitudes towards gender and family roles					
Different gender suitability for professions	0.00	1.00	-1.90	1.28	8,868
Different gender duties in the home	0.00	1.00	-1.29	3.55	18,008
Gender use of technical devices	0.00	1.00	-1.06	2.52	8,859
Attitude towards marriage	0.00	1.00	-1.35	1.35	14,943

Notes: Continuation on next page

Descriptive statistics (continued)

	Mean	Std. Dev.	Min.	Max.	Obs.
Family and labor-market outcomes					
Currently married	0.60	0.49	0.00	1.00	56,673
Number of children	1.38	1.25	0.00	12.00	52,668
Labor-force participation	0.82	0.38	0.00	1.00	58,168
Employment	0.78	0.41	0.00	1.00	58,168
Working hours	35.56	14.89	0.00	120.00	45,781
Earnings	7.14	0.90	0.00	11.61	44,935
Educational outcomes					
Years of education	12.96	2.83	6.00	25.00	42,772
Abitur	0.38	0.49	0.00	1.00	52,283
Age of first employment	21.11	3.88	1.33	65.25	38,985
Controls					
Female	0.51	0.50	0.00	1.00	58,703
Migration status	0.05	0.22	0.00	1.00	58,703
Mother's education					
Basic (Hauptschulabschluss or less)	0.61	0.49	0.00	1.00	58,703
Medium (Realschulabschluss)	0.18	0.39	0.00	1.00	58,703
High (Abitur or more)	0.09	0.29	0.00	1.00	58,703
Father's education					
Basic (Hauptschulabschluss or less)	0.57	0.50	0.00	1.00	58,703
Medium (Realschulabschluss)	0.13	0.34	0.00	1.00	58,703
High (Abitur or more)	0.16	0.36	0.00	1.00	58,703
NEPS	0.21	0.41	0.00	1.00	58,703
ALLBUS	0.27	0.44	0.00	1.00	58,703
SOEP	0.52	0.50	0.00	1.00	58,703

Notes: Descriptive statistics. The sums of the category means of mother's and father's education, respectively, do not add up to one because missing values are set to zero, defining a separate binary explanatory variable that accounts for the missing values. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.3: Effects of abolishing compulsory religious education on religiosity and religious actions

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.071	-0.046	-0.066	-0.081
	(0.018)	(0.101)	(0.020)	(0.009)
	[0.061]	[0.136]	[0.022]	[0.066]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,688	13,276	42,776	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Appendix Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.4: Time-varying treatment effects on religious outcomes

	Religiosity		Prayer		Church-going		Affiliation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform	-0.072	0.017	-0.045	0.037	-0.049	0.005	-0.087	-0.054
	(0.031)	(0.593)	(0.129)	(0.159)	(0.063)	(0.906)	(0.005)	(0.034)
	[0.149]	[0.733]	[0.214]	[0.209]	[0.075]	[0.925]	[0.052]	[0.068]
Years relative to reform	0.000	0.002	-0.000	0.001	-0.007	-0.006	0.003	0.003
	(0.942)	(0.611)	(0.821)	(0.660)	(0.007)	(0.015)	(0.135)	(0.051)
	[0.941]	[0.731]	[0.822]	[0.715]	[0.284]	[0.328]	[0.231]	[0.149]
Reform x		-0.013		-0.012		-0.007		-0.004
Years relative to reform		(0.001)		(0.001)		(0.161)		(0.129)
		[0.105]		[0.035]		[0.480]		[0.288]
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,688	15,688	13,276	13,276	42,776	42,776	45,925	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Appendix Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.5: Heterogeneous treatment effects on religious outcomes

	Religiosity		Prayer		Church-going		Affiliation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Gender	Females	Males	Females	Males	Females	Males	Females	Males
Reform	-0.067	-0.073	-0.085	-0.007	-0.039	-0.097	-0.075	-0.085
	(0.024)	(0.060)	(0.057)	(0.835)	(0.251)	(0.009)	(0.012)	(0.025)
	[0.033]	[0.237]	[0.114]	[0.841]	[0.179]	[0.037]	[0.112]	[0.094]
Panel B: Area	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Reform	-0.067	-0.123	-0.100	-0.024	–	–	-0.064	-0.040
	(0.038)	(0.071)	(0.037)	(0.615)			(0.131)	(0.670)
	[0.007]	[0.102]	[0.034]	[0.572]			[0.196]	[0.695]
Panel C: Area	Catholic	Protestant	Catholic	Protestant	Catholic	Protestant	Catholic	Protestant
Reform	-0.157	-0.016	-0.124	-0.041	–	–	-0.211	0.064
	(0.009)	(0.687)	(0.004)	(0.482)			(0.001)	(0.285)
	[0.021]	[0.655]	[0.015]	[0.468]			[0.017]	[0.317]
Panel D: Parents	Catholic	Protestant	Catholic	Protestant	Catholic	Protestant	Catholic	Protestant
Reform	–	–	–	–	-0.071	0.004	-0.077	-0.113
					(0.199)	(0.904)	(0.044)	(0.047)
					[0.324]	[0.903]	[0.199]	[0.120]

Notes: Each cell reports the coefficient on reform treatment from a separate regression. All regressions include state and birth-year fixed effects and controls. Dependent variables indicated in column headers. All dependent variables are standardized (see Appendix Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Urban area if county has more than 100,000 inhabitants; rural otherwise (available only for RemoteNEPS). Catholic area if number of Catholics over sum of Protestants and Catholics in county is larger than 0.5; Protestant area otherwise (available only for RemoteNEPS). Catholic/Protestant parents if both parents are Catholic/Protestant (available only for ALLBUS and SOEP). Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap byRoodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.6: Effects on ethical-value and political-value outcomes

Panel A: Ethical-value outcomes

	Reciprocity	Trust	Risk-taking	Volunteering	Life satisfaction
	(1)	(2)	(3)	(4)	(5)
Reform	0.006	0.007	0.008	0.007	-0.014
	(0.734)	(0.780)	(0.636)	(0.681)	(0.478)
	[0.748]	[0.816]	[0.748]	[0.792]	[0.682]
State and birth-year fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	21,150	37,070	35,556	37,971	48,177

Panel B: Political-value outcomes

	Interest in politics	Politics too complicated	Satisfaction with democracy	Political spectrum: right	Vote in election	Vote left	Vote extreme
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reform	0.010	0.017	0.001	-0.021	0.011	-0.016	-0.004
	(0.530)	(0.675)	(0.980)	(0.195)	(0.070)	(0.245)	(0.477)
	[0.603]	[0.718]	[0.992]	[0.249]	[0.128]	[0.404]	[0.485]
State and birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	52,970	9,160	14,519	40,161	32,133	27,088	27,100

Notes: Dependent variables indicated in column headers. Dependent variables (see Appendix Table A2.2 for details): panel A: columns (1) – (3), (5): standardized; column (4): indicator variable; panel B: columns (1) – (4): standardized; columns (5) – (7): indicator variable. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.7: Effects on family and labor-market outcomes

Panel A: Attitudes towards gender and family roles

	Different gender suitability for professions	Different gender duties in the home	Gender use of technical devices	Attitude towards marriage
	(1)	(2)	(3)	(4)
Reform	-0.084	-0.035	-0.061	-0.117
	(0.084)	(0.371)	(0.005)	(0.002)
	[0.183]	[0.452]	[0.044]	[0.044]
State and birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	8,868	18,008	8,859	14,943

Panel B: Family and labor-market outcomes

	Married	Number of children	Labor-force participation	Employment	Working hours	Earnings
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.015	-0.088	0.015	0.023	0.590	0.053
	(0.114)	(0.006)	(0.002)	(0.000)	(0.095)	(0.032)
	[0.074]	[0.031]	[0.036]	[0.002]	[0.168]	[0.057]
State and birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,673	52,668	58,168	58,168	45,781	44,935

Notes: Dependent variables indicated in column headers. Dependent variables (see Appendix Table A2.2 for details): all dependent variables in panel A are standardized; panel B: columns (1), (3), (4): indicator variable; columns (2), (5): numbers; column (6): log earnings. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.8: Effects on educational outcomes

	Years of education	Abitur	Age at first employment
	(1)	(2)	(3)
Reform	0.032	-0.023	0.018
	(0.670)	(0.075)	(0.866)
	[0.730]	[0.226]	[0.899]
State fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	42,772	52,283	38,985

Notes: Dependent variables indicated in column headers. Dependent variables (see Appendix Table A2.2 for details): column (1), (3): number; column (2): indicator variable. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table 2.9: Effects on religious outcomes: Border specification with county-pair fixed effects

	Religiosity	Prayer	Affiliation
	(1)	(2)	(3)
Reform	-0.162	-0.169	0.006
	(0.022)	(0.063)	(0.883)
	[0.007]	[0.036]	[0.877]
State fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	3,070	2,648	3,072

Notes: Dependent variables indicated in column headers (church-going not covered in NEPS data). All dependent variables are standardized (see Appendix Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey, survey-year fixed effects, and bordering-county-pair fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6.

Appendix

A2.1 Data Appendix

This appendix provides additional detail on the three individual-level datasets and their preparation and describes how we merge them for our analysis.

National Education Panel Study (NEPS)

The National Education Panel Study (NEPS) is a large-scale longitudinal survey capturing educational biographies of individuals in Germany (Blossfeld and Maurice, 2011). It focuses not only on competencies, educational processes, educational decisions, and returns to education throughout the life span of individuals, but also covers a wide range of other topics including several questions on religiosity. NEPS has six different "starting cohorts", from newborns to adults, which are then followed through their lives.[1]

We use Starting Cohort 6 which covers the educational and professional careers of a representative sample of adults with a special focus on adult education and lifelong learning. The survey was first administered in 2007/2008 with seven follow-up waves until 2015/2016.

Whenever a variable of an individual is measured in multiple waves, we use its most recent non-missing value. The data cover detailed retrospective questions on the educational biographies of respondents including the state and year of primary school entry, which we use to link the status of compulsory religious education for this state-year combination. Whenever the state of the primary school location is not available, we use the state of residence of the individual in the primary school entry year instead if available. Whenever the year of primary school entry is not available, we use the year of secondary school entry minus four, if available, given the default duration of primary school equals four years in Germany.

We keep individuals in the sample who provide information about their state and year of primary school entry, as well as about basic control variables (gender and migration background). We further require that the individuals entered primary school after 1949 and before 2005 in a West German state. The resulting sample consists of 12,281 individuals.

We keep individuals in the sample who provide information about their state and year of primary school entry, as well as about basic control variables (gender and migration background). We further require that the individuals entered primary school after 1949 and before 2005 in a West German state.[2] The resulting sample consists of 12,281 individuals.

[1] One "starting cohort" contains many birth cohorts. The Starting Cohort 6, which we use in our analysis, includes birth cohorts from 1944 to 1988.

[2] For Baden-Württemberg and Saarland, we only keep individuals in the sample who entered primary school after 1952 and 1956, respectively, as the legal status of religious education was not defined or cannot be retrieved from legal documents for the previous years (Helbig and Nikolai, 2015)).

Regarding religious outcome variables, NEPS contains our main outcome variable religiosity as well as information on personal prayer and religious affiliation. Church-going is not included. NEPS also contains most variables from the other outcome variable groups (Table A2.2). Compared to ALLBUS and SOEP, gender role attitudes are particularly well covered.

Regarding control variables, NEPS contains information on gender, migration status, father's and mother's education, and the survey year. Missing values of father's and mother's education are set to zero, and a separate binary explanatory variable is introduced that accounts for the missing values. Given our approach to use the most recent available information per individual and variable, we store the survey year of an individual separately for each outcome variable and use it accordingly as outcome-specific control variable in the regression analyses. In contrast to ALLBUS and SOEP, information on the religious affiliation of the parents is not available in NEPS.

To access fine-grained geographical information below the state level, we make use of RemoteNEPS, the technology that enables remote data processing of sensitive information. RemoteNEPS provides the county identifier of an individual's primary school location, which we merge to administrative data about the county structure (rural vs. urban, Catholic vs. Protestant). In addition, we can use this information to implement our border specification of individuals going to school in counties neighboring each other across state borders (and including county-pair fixed effects).[3]

German General Social Survey (ALLBUS)

The German General Social Survey (ALLBUS) is a biennial cross-sectional survey that monitors societal change by interviewing a nationally representative sample of adults in Germany since 1980 (GESIS, 2019). It provides a picture of the attitudes, behaviors, and social structure of the population in Germany. We use the ALLBUS Cumulation that combines 20 waves from 1980 to 2016.[4] The ALLBUS Cumulation contains all variables from the twenty waves that are elicited in at least two waves. Unlike NEPS and SOEP, the cross-sectional data structure of ALLBUS implies that each individual is observed only once.

The data contain information on the state a respondent lived in during childhood, which we assume is the primary school entry state. If this information is not available, we assume that the respondent entered primary school in her state of birth. Unlike NEPS, ALLBUS does not elicit the year of primary school entry. We assume that respondents entered primary school six years after their birth year, given that most students enter primary school at the age of six in Germany. We then merge the state-level data on compulsory religious education to the thus defined state and year of primary school entry of each individual.

[3] SOEP also has a remote feature which would allow to access information on the county of residence, but not the county of schooling. In addition, it would be infeasible to merge other datasets with RemoteNEPS.

[4] Beyond the biennial survey pattern, there was one additional wave administered in 1991.

We keep all individuals in the sample who provide the variables to approximate the state and year of primary school entry as well as basic control variables, and who entered primary school after 1949 and before 2005 in a West German state. The overall sample size equals 15,924 individuals. However, the number of observations varies substantially between variables, as not all questions were asked in all waves.

ALLBUS is the only dataset that contains all of our four religious outcome variables – religiosity, prayer, church-going, and affiliation. It is also comprehensive with regards to the other outcome variables, with the exception that it only covers two variables on attitudes towards gender and family roles (different gender duties in the home and attitudes towards marriage, see Table A2.2). ALLBUS contains the same basic control variables as NEPS. In addition, it provides information on the religion of the mother and father for a subset of individuals. We apply the same approach to address missing values described above for NEPS to ALLBUS and SOEP.

German Socio-Economic Panel (SOEP)

The German Socio-Economic Panel (SOEP) is a representative longitudinal survey of private households and individuals in Germany. It covers many topics including household composition, occupational biographies, employment, earnings, health, and satisfaction. We employ the SOEP Core 1984-2017 (v.34) which follows individuals since 1984 and has been repeatedly supplemented with new samples to account for changes that took place in the German society, such as samples of migrants and refugees (Goebel et al., 2019). Analogous to NEPS, we use the most recent available non-missing value of a variable for each individual.

To approximate the state and year of primary school entry, we assume that individuals entered primary school in the state of their last school attendance, which is elicited in SOEP for a subset of respondents. For the other respondents, we assume that they entered primary school in their state of birth. As in ALLBUS, we assume that individuals entered primary school six years after their birth and accordingly merge status information on compulsory religious education.

We again keep all individuals in the sample who provide the variables to approximate the state and year of primary school entry as well as basic control variables, and who entered primary school after 1949 and before 2005 in a West German state. The resulting sample size equals 30,498 individuals.

SOEP contains two of the four religious outcome variables (church-going and religious affiliation) and two of the four variables measuring attitudes towards gender and family roles (different gender duties in the home and attitudes towards marriage, see Table A2.2). However, SOEP provides a comprehensive set of other outcomes, with a special focus on labor-market, educational, and ethical-value outcomes. In terms of control variables, SOEP is comparable

to ALLBUS: In addition to the main control variables, it also contains information about the religion(s) of the mother and father for a subset of individuals.

Merging the three Datasets

NEPS, ALLBUS, and SOEP are collected independently from each other. Hence, their data structure and variables are not aligned. To merge the three datasets, we start by evaluating the questionnaires of the three datasets and select only variables for the merging procedure whose question wordings in the questionnaires are directly comparable.

For each selected variable, we recode the answer categories in each dataset to be directly comparable across datasets. This implies standardization in most cases, but occasionally also requires the recoding of variables to analogous dummy or categorical variables. Table A2.2 provides a list of the precise wording and number of answer categories for all outcome variables for each of the three datasets.

For example, our main outcome variable religiosity in NEPS is phrased, “Faith and religion are part of everyday life for some people. What about you? Regardless of whether you belong to a religious community, how religious would you say you are?” There are four answer categories, “Not at all religious”, “Slightly non-religious”, “Slightly religious”, and “Very religious”. In ALLBUS, the question on religiosity is phrased, “Would you describe yourself as more religious or more not religious? We have a scale for this. Where would you place yourself on this scale?” The ten answer categories range from “not religious” to “religious”. In SOEP, there is no question on religiosity. Because of the different answer categories in NEPS and ALLBUS, both religiosity variables are standardized before being merged together.

Other variables also required re-coding of answer categories before standardization such that an increase in the variable implies a change in the same direction across datasets. For example, an increase in the raw variable on personal prayer in NEPS implies a decrease in the propensity to pray, whereas an increase in the corresponding raw variable in ALLBUS implies an increase in the propensity to pray. Throughout the paper, all answer categories are ordered before standardization such that an increase in the variable implies an increase in religiosity. The same is true for conservative attitudes towards gender and family roles.

Before merging the datasets, we create three dummy variables, one for each dataset, to indicate the respective data source. Finally, we order all variables analogously in the three datasets and then append NEPS with ALLBUS and SOEP.

A2.2 Diagnostics of the Two-way Fixed Effects Specification

This appendix reports two diagnostic tests of the two-way fixed effects specification that complement the results of the Callaway and Sant'Anna (2021) estimator reported in the main text.

Diagnostics by de Chaisemartin and D'Haultfœuille (2020)

The diagnostic test by Chaisemartin and D'Haultfœuille (2020) is based on the observation that the estimate derived from a two-way fixed effects difference-in-differences estimation under the common trend assumption is a weighted sum of the average treatment effect in each group and period. Heterogeneity in treatment effects can lead to negative weights attached to specific group-period estimates. When estimating the weights of the group-period clusters in our setting, 46 of the 216 ATTs receive a negative weight, which sum to -0.070. Investigation indicates that negative weights are particularly frequent in estimates involving the two always-treated states in our setting, Berlin and Bremen, which effectively had adopted the reform by the time our sample starts in 1950.

When conducting the analysis without the two always-treated states, only five of the 125 ATTs receive a negative weight, which sum to only -0.004. Reassuringly, estimates of the treatment effects on all religious outcomes in our main specification are qualitatively unaffected when excluding Berlin and Bremen (see Appendix Table A2.16).

Decomposition by Goodman-Bacon (2021)

In addition, we perform the Goodman-Bacon (2021) decomposition to display potential heterogeneity in the estimated effect components and clarify which relationships and groups matter most. Specifically, we analyze whether our main result holds in a subset of effect components that is immune to biases from negative weighting. To implement the analysis, we collapse data to means of state-cohort cells. To create a balanced panel, we drop observations in cohorts before 1949 or after 1991, which implies deletion of 21 percent of all state-cohort cells.

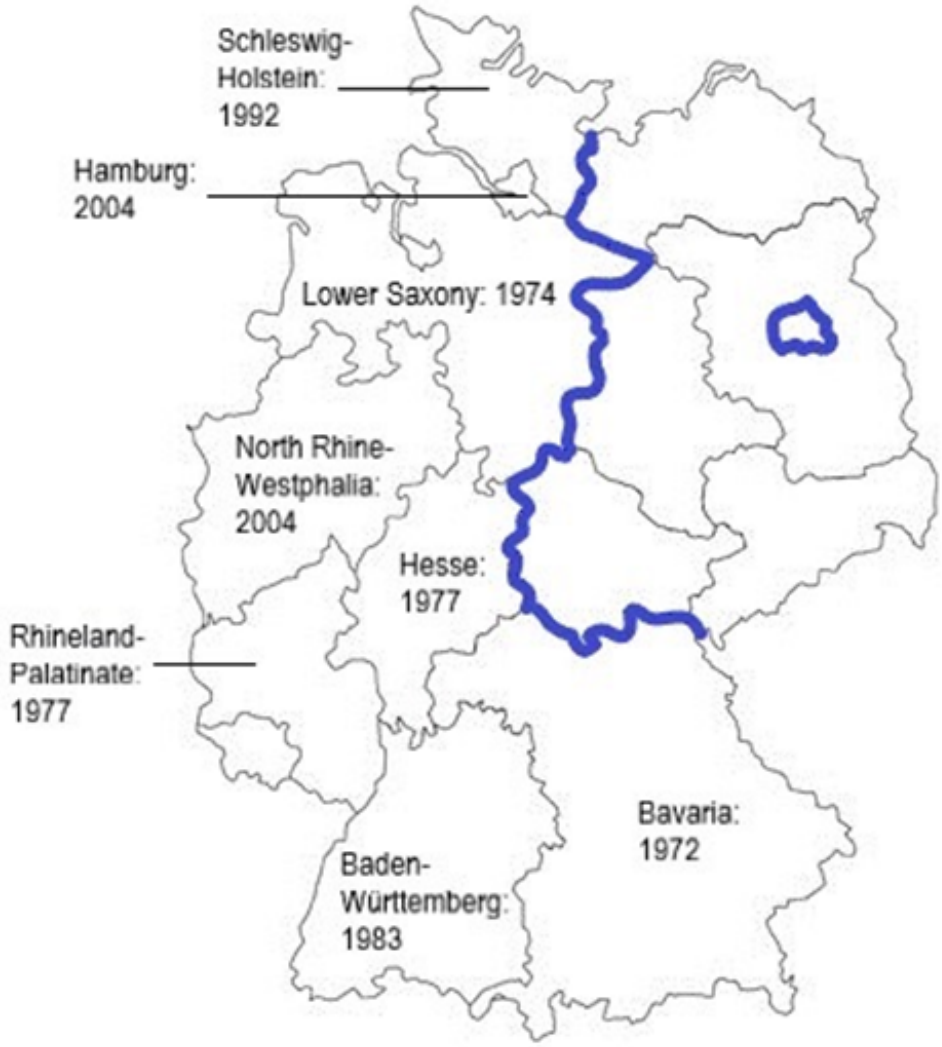
The graphs contained in Appendix Tables A2.17-A2.20 show scatterplots of two-by-two difference-in-differences estimates and their associated weights for the four measures of religious outcomes. The figures depict three types of two-group/two-period comparisons that differ by control group: (1) timing groups, i.e., groups whose treatment at different times serves as each other's control groups in two ways: those treated later serve as the control group for an earlier treatment group and those treated earlier serve as the control group for the later group; (2) always treated, where a group treated prior to the start of the analysis serves as the control group; and (3) never treated, where a group which never receives the treatment serves as the control group. In our setting, the two always-treated states are Berlin and Bremen.

There is one never-treated state that never adopted the reform (Saarland). All other West German states adopted the reform within our estimation sample from 1950 to 2004.

The difference-in-differences estimators derived from the Goodman-Bacon (2021) decomposition, shown in the first line of Appendix Tables A2.17-A2.20, are similar to the results of our main specification. The estimator is in fact larger in absolute terms for three of the four religious outcomes, and only slightly smaller for religious affiliation. The overall effect of the reform on religiosity is -0.129 (compared to -0.071 in our main specification of Table 2.3). Across all four religious outcomes, the never vs. timing comparison receives the largest weight. This comparison is immune to biases from time-varying treatment effects and, reassuringly, displays a negative effect in all four decompositions. Overall, results of the diagnostic tests thus indicate that our findings are not driven by a setting that would give rise to negative weights.

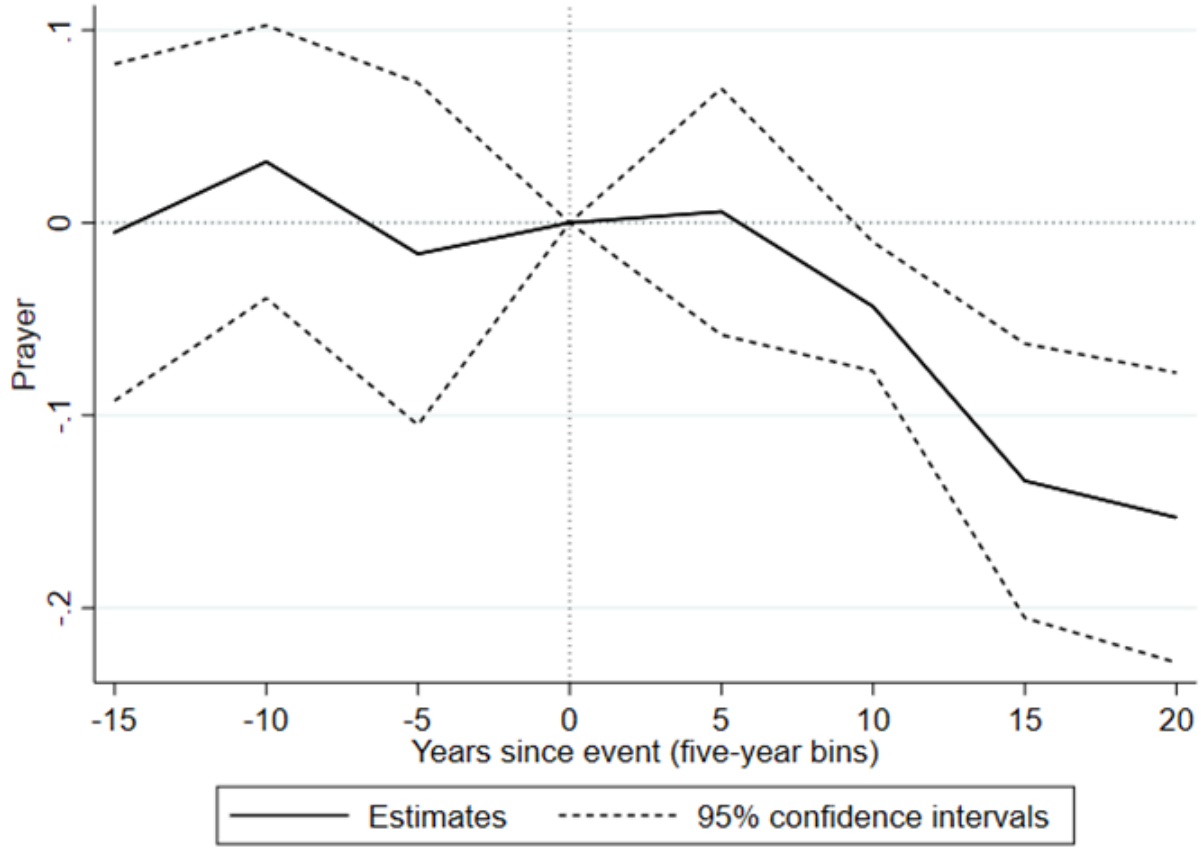
A2.3 Appendix Figures and Tables

Figure A2.1: Religious education reforms in West German states



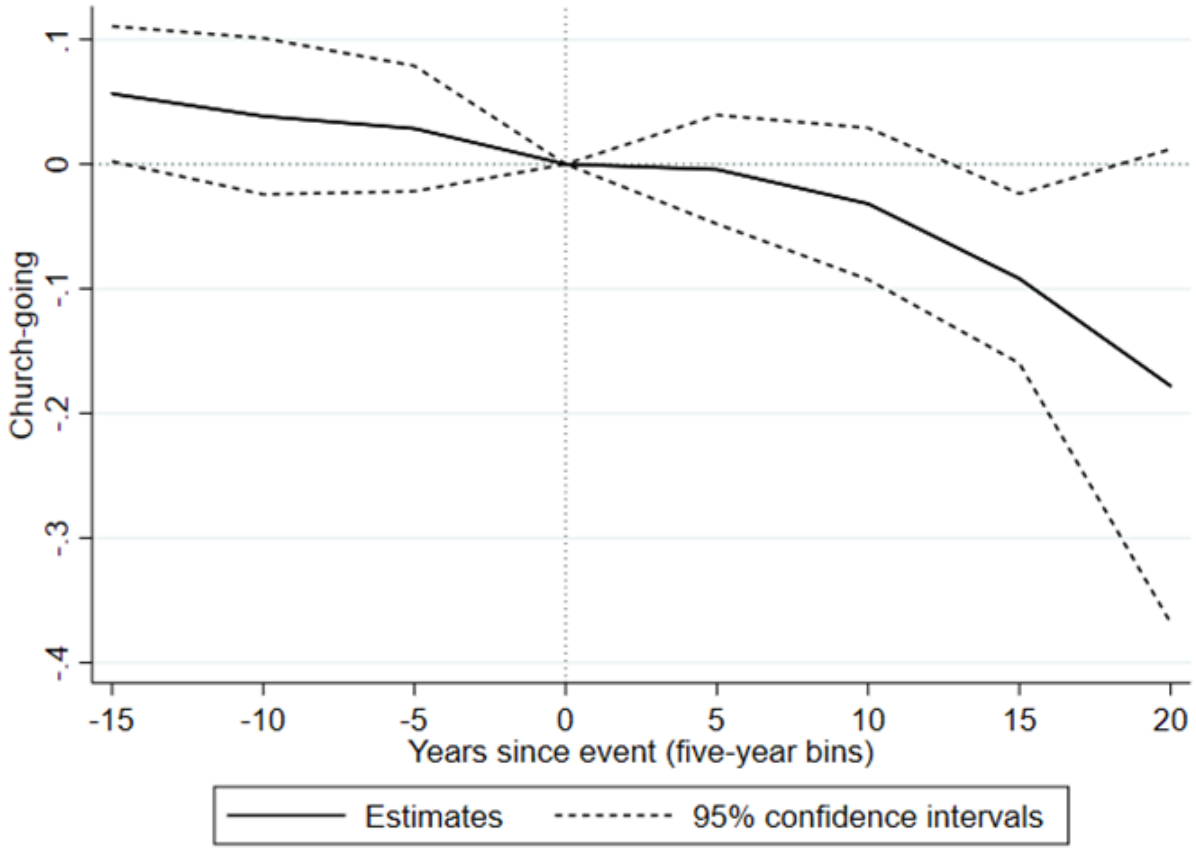
Notes: Map displays years of the abolishment of compulsory religious education of West German states.

Figure A2.2: Non-parametric event-study estimates of effect on personal prayer



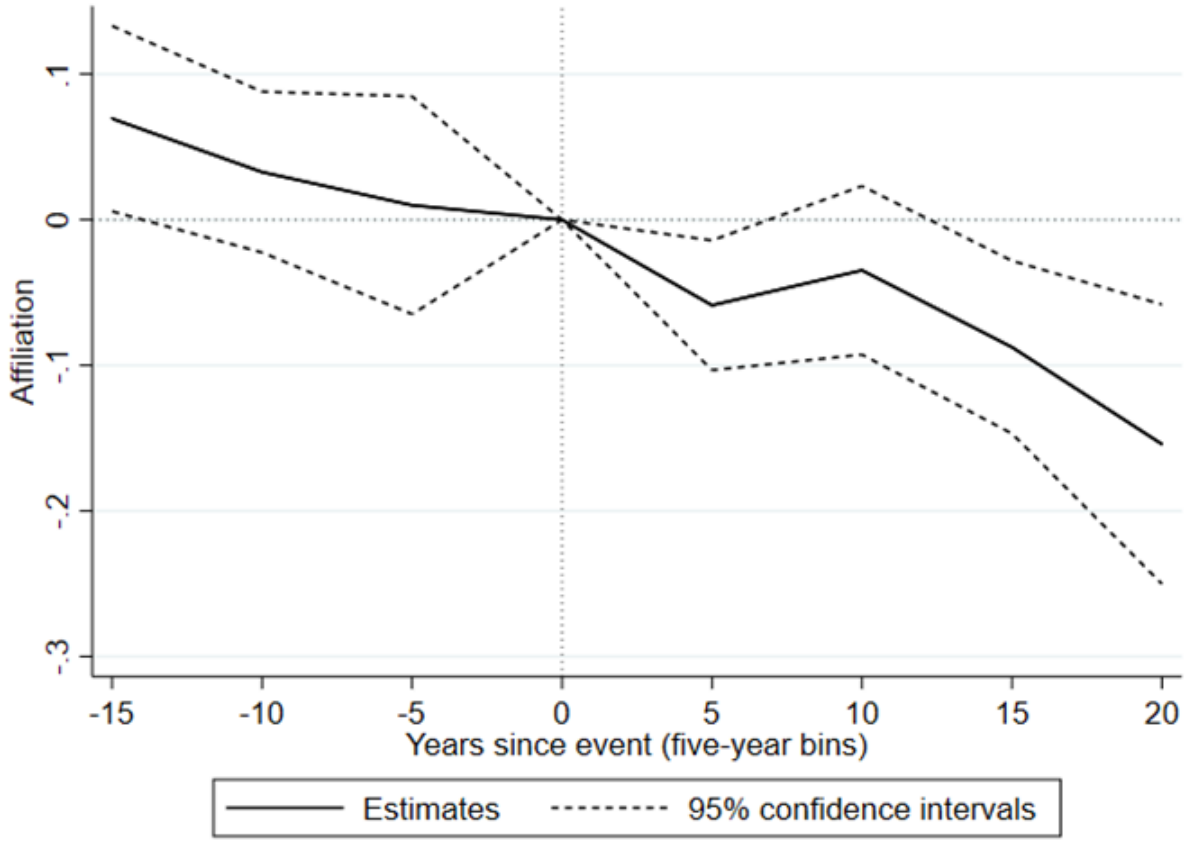
Notes: Coefficients from non-parametric event-study regressions and their 95 percent confidence intervals. Dependent variable: personal prayer (standardized, based on 7-point-scale NEPS question "How often do you pray?" and the same 11-point-scale ALLBUS question). Numbers on horizontal axis refer to final year of respective five-year bins; i.e., 0 = last five years prior to treatment (excluded category), 5 = first five years of treatment. Inference: Standard clustering at state level. The p-values of omnibus hypothesis tests of zero pre- and post-event effects are 0.588 and 0.003, respectively. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Figure A2.3: Non-parametric event-study estimates of effect on church-going



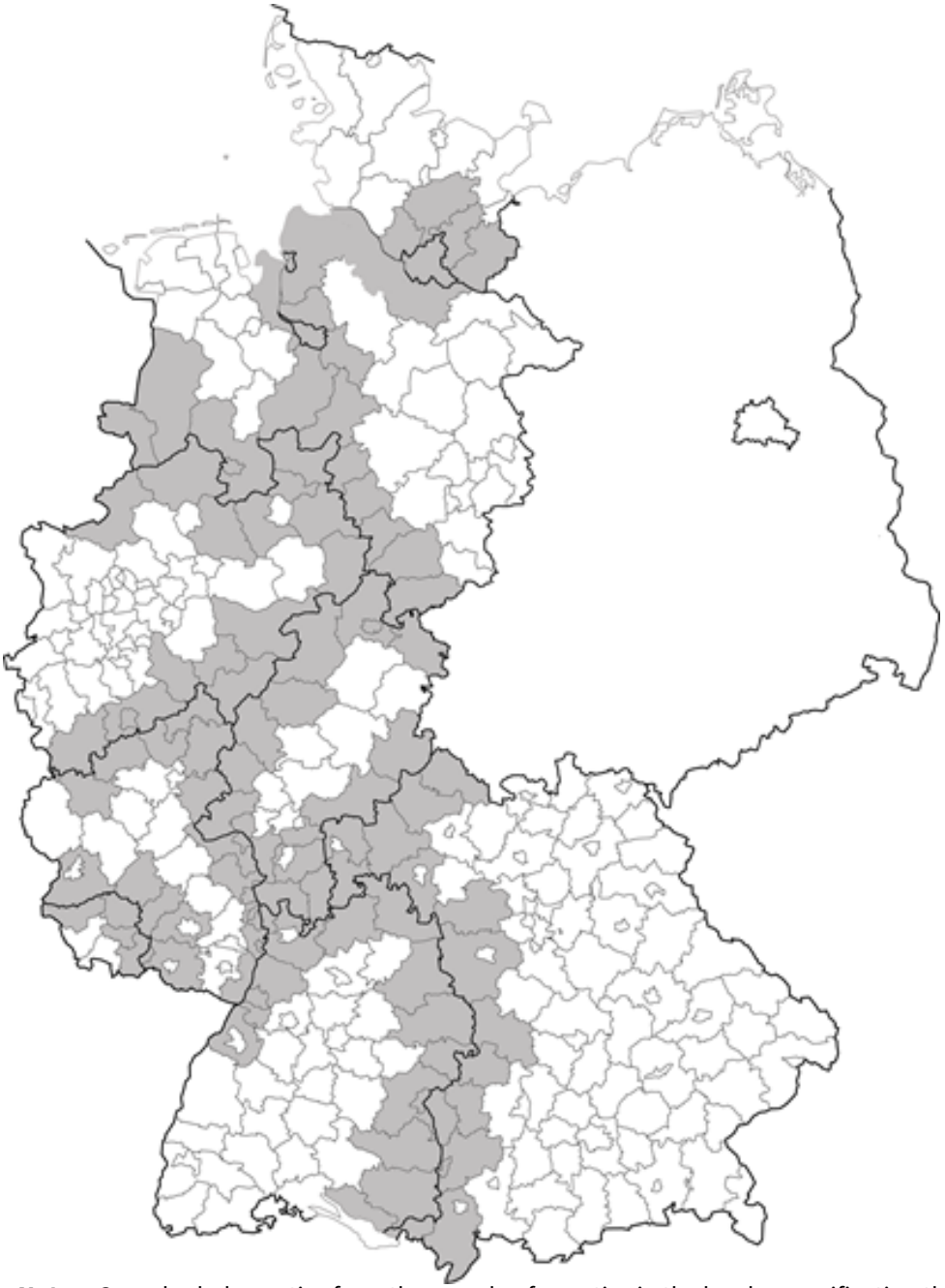
Notes: Coefficients from non-parametric event-study regressions and their 95 percent confidence intervals. Dependent variable: church-going (standardized, based on 6-point-scale ALLBUS question "As a rule, how often do you go to church?" and 4-point-scale SOEP question "Which of the following activities do you take part in during your free time? Attending church, religious events"). Numbers on horizontal axis refer to final year of respective five-year bins; i.e., 0 = last five years prior to treatment (excluded category), 5 = first five years of treatment. Inference: Standard clustering at state level. The p-values of omnibus hypothesis tests of zero pre- and post-event effects are 0.139 and 0.087, respectively. Data sources: German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Figure A2.4: Non-parametric event-study estimates of effect on religious affiliation



Notes: Coefficients from non-parametric event-study regressions and their 95 percent confidence intervals. Dependent variable: religious affiliation (standardized, based on 6-point-scale ALLBUS question "Which religion do you belong to?" and 11-point scale SOEP question "Do you belong to a church, religious community or faith?"). Numbers on horizontal axis refer to final year of respective five-year bins; i.e., 0 = last five years prior to treatment (excluded category), 5 = first five years of treatment. Inference: Standard clustering at state level. The p-values of omnibus hypothesis tests of zero pre- and post-event effects are 0.052 and 0.020, respectively. Data sources: German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Figure A2.5: Counties in the border specification



Notes: Grey shaded counties form the sample of counties in the border specification that are directly at the border to another state. Thick and thin grey lines represent state and county borders, respectively.

Table A2.1: Curricula before and after the reform: The case of Bavaria

Syllabus of Catholic religious education, 1967	Syllabus of Catholic religious education, 1979	Syllabus of Ethics, 1986
Religious education in the classroom is only supposed to be a part of a bigger all-encompassing religious education, the catechesis. (p. 666, Part IV 4.) Catechesis is defined as the instruction and indoctrination of the practices and beliefs of the Catholic Church. (p. 663, Part III 1.) (p. 666, Part IV 5.) Religious education teachers are warned from adopting general pedagogical values that are taught in school since they are based on anthropocentric pedagogic which is not compliant with the nature of catechesis. (p. 667, Part V)	Content-wise religious education is supposed to be seen as relating to catechesis, but the organization and shape of religious education as a school subject underlies the educational mandate of the state. (p. 102, 1.)	The syllabus discusses ethical questions that are important in life in our time and covers them more and more in depth in every grade. Open and focused discussion is important and should be encouraged by the teacher. In discussion students are supposed to formulate their own findings and use them to explore ethical solutions and ways of acting for different situations in life. Discussions of ethical questions are supposed to be based on concrete situations that are drawn from real life.
Religious education is to be: - Centered around reality not some abstract concept - Integrated in school and not separate from all other courses - Focused on Jesus Christ and bringing across the "basic truth" and "central message" of the catholic faith and the bible - Dialogical structure of education as interpersonal process which is based in encounter/connection between god and man - Application of living with faith (pp. 667-668, Part V)	**Religious education is to be:** - Very clearly structured with set topics, content, methods and controls - Learning targets and content of lessons are mandatory and must be covered as laid out in syllabus - Interactive, events and field trips should be used to enhance learning experience and make it more connected to life (pp. 103-106)	

Notes: Continuation on next page

Curricula before and after the reform: The case of Bavaria (continued)

Syllabus of Catholic religious education, 1967	Syllabus of Catholic religious education, 1979	Syllabus of Ethics, 1986
Main goals of religious education:	**Main goals of religious education:**	**Main goals of ethics:**
- Introduction and instruction of prayer as the central way of self-disclosure to god	- Religious education is supposed to enable responsible thinking and behavior based on religion and faith	- Guide students towards responsible actions in their personal life and in society
- Guide to having the church in one's life	- Reflect on and question the purpose of human life and the world	- Show the commonalities of general ethical values and Christian values
- Guide to dealing with the unfaith of one's environment	- Break up pretended faith and thoughtless unfaith. In doing so helping to prevent a degeneration of pluralism into "passive indifference".	- Teaching tolerance towards others
- Formation of one's conscience	- Aid faithful student to be more actively connected to religion, aid the "searching" student in finding the answers of the church to his questions, give the unfaithful student opportunity to become clearer in his viewpoint or change it (pp. 102-103)	
- Gender education must be done with help of parents (pp. 669-672, Part VI)		
Syllabus in grades 5-10:	**Syllabus in grades 5-10:**	**Syllabus in grades 5-10:**
- Grade 5-6: kids' development peaks, they are increasingly able to think critically	- Grade 5: students are supposed to recognize faith and religion as something to guide them (pp. 107-114)	- The two main topics are "Man and his/her personal life" and "Man in a society with others"
- Grade 7: puberty causes "a crisis" and kids change their attitude about what's important and who their role models are, also "sexual impulses disturb the young adult"	- Grade 6: As they approach the end of their childhood students are supposed to capture and open themselves up to the guiding power of the Christian faith (pp. 115-123)	- Every grade works on both topics so that both are only fully covered at the end of ninth grade.
- Grades 8-9 are supposed to cover the current themes in the church from a historical standpoint	- Grade 7: in a time of personal insecurity students are supposed to discover how faith can help solve their own problems and difficulties (pp. 123-131)	- Subtopics in "Man and his/her personal life": time-management, good deeds, seeing beauty, independent learning, meaningful free-time activities, making decisions, social impact of work …
- Grade 9 is supposed to help young adults answer important questions in their life	- Grade 8: Amidst puberty students are supposed to experiment with the Christian way of life and consider it a serious possibility in shaping their own life (pp. 132-138)	- Subtopics in "Man in a society with others": being accepted, ending conflict, behavior towards strangers/foreigners, causes for prejudice, respecting freedom of opinion, meaning of authority, meaning of guilt, dealing with guilt, …
- Despite its big advantages religious education in the classroom is limited through the compulsory atmosphere in school and should be complemented with "religious community days" (pp. 679-689, Part Lehrpläne)	- Grade 9: at the end of the first period in their life students are supposed to perceive faith as life-improving and life as being open for faith (page 139-144)	
Syllabus is structured by giving one or two topics per grade. Topics are given without specific guidance on how to teach these topics.	Syllabus states up to eight topics, for every grade, all of which are explained in how they are to be taught and what they should encompass.	Syllabus is divided into two main topics which are both discussed fifth through ninth/tenth grade. In each grade, different subtopics of the two main topics are discussed without specific guidance how these topics are to be taught.

Notes: Own depiction based on the respective curricula as published in the Amtsblatt des bayerischen Staatsministeriums für Unterricht und Kultus.

Table A2.2: Outcome measures derived from the three datasets

	NEPS (1)	ALLBUS (2)	SOEP (3)
Religious outcomes			
Religiosity (s)	Regardless of whether you belong to a religious community, how religious would you say are you? (4)	Would you describe yourself as more religious or more not religious? (10)	–
Prayer (s)	How often do you pray? (7)	How often do you pray? (11)	–
Church-going (s)	–	As a rule, how often do you go to church? (6)	Which of the following activities do you take part in during your free time? Attending church, religious events (4)
Affiliation (s)	Do you belong to a faith or religion? (2)	Which religion do you belong to? (6)	Do you belong to a church, religious community or faith? (11)
Ethical-value outcomes			
Reciprocity (s)	–	I go out of my way to help somebody who has helped me before. (7)	I make particular effort to help someone who has previously helped me. (7)
Trust (s)	I trust other people easily, I believe in the goodness in people (5)	Some people think that most people can be trusted. Others think that one can't be careful enough when dealing with other people. What do you think? (4)	On the whole trust people (4)
Risk-taking (s)	How do you assess yourself: Are you generally willing to take risks or do you try to avoid risks? Please respond on a scale from 0 to 10. '0' indicates that you are not willing to take risks while '10' means that you are very much willing to take risks. You can use the numbers in between to stagger your answer. (11)	–	How do you rate yourself personally? In general, are you someone who is ready to take risks or do you try to avoid risks? (11)
Volunteering (d)	Have you ever been actively involved in clubs, organizations, initiatives or self-help groups before? (3)	Please tell me here, too, how often you do the following in your leisure time: Voluntary work in clubs, associations or community services (5)	Which of the following activities do you take part in during your free time? Please check off how often you do each activity: at least once a week, at least once a month, less often, never. Volunteer work in clubs or social services (4)

Notes: Continuation on next page

Outcome measures derived from the three datasets (continued)

	NEPS (1)	ALLBUS (2)	SOEP (3)
Life satisfaction (s)	I would like to begin by asking you a few questions about your current satisfaction with different aspects of your life. Please answer on a scale of 0 to 10. '0' means that you are entirely unsatisfied, '10' means that you are entirely satisfied. You can grade your assessment using the numbers in between. In general, how satisfied are you currently with your life? (11)	And now another general question. How satisfied are you – all in all – with your life at the moment? (11)	In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered? (11)
Political-value outcomes			
Interest in politics (s)	How much are interested in politics? Are you very interested, rather interested, little interested or not interested at all? (4)	How interested in politics are you? Very strongly, strongly, middling, very little, or not at all? (5)	Generally speaking, how interested are you in politics? (4)
Politics too complicated (s)	How often do politics seems so complicated to you that you don't really understand what it's all about? (5)	On this list, there are a number of opinions one can hear now and then. For each opinion, please tell me if you: completely agree, tend to agree, tend not to agree, or completely disagree: Politics is so complicated that somebody like me can't understand what's going on at all. (4)	–
Satisfaction with democracy (s)	–	Let's turn to democracy in Germany: Generally speaking, how satisfied are you with democracy as practiced in Germany? (6)	Satisfaction with democracy in Germany (11)
Political spectrum: right (s)	In politics you sometimes talk about 'left' and 'right'. Where on a scale from 0 to 10 would you grade yourself, if 0 is left and 10 is right? (11)	Many people use the terms "left" and "right" to describe differing political views. Here we have a scale that runs from left to right. If you think of your own political views, where would you place them on this scale? (10)	In politics people often talk about "left" and "right" when it comes to characterize different political attitudes. If you think about your own political views: Where would you place yours? (11)
Vote in election (d)	Some people do not vote nowadays for various reasons. What about you? Did you vote during the last Bundestag election? (2)	Did you vote in last federal election? (2)	Attendance Bundestag election 2013 (2)
Vote left (d)	If Bundestag elections were to be held tomorrow, which party would you give your second vote to? (8)	If there was a federal election next Sunday, which party would you vote for with your second vote? (10)	And how was it at the last general election (Bundestagswahl) on September 22, 2013? Which party did you vote for? (9)
Vote extreme (d)	If Bundestag elections were to be held tomorrow, which party would you give your second vote to? (8)	If there was a federal election next Sunday, which party would you vote for with your second vote? (10)	And how was it at the last general election (Bundestagswahl) on September 22, 2013? Which party did you vote for? (9)

Notes: Continuation on next page

Outcome measures derived from the three datasets (continued)

	NEPS (1)	ALLBUS (2)	SOEP (3)
Attitudes towards gender and family roles			
Different gender suitability for professions (s)	Men are better suited for certain professions than women. Do you completely disagree, somewhat disagree, somewhat agree or agree completely? (4)	–	–
Different gender duties in the home (s)	Men and women should have the same duties in the home (4)	How do you and your partner share these activities in your household? Who does what? Cleaning the house/flat (6)	Men involved in housework (4)
Gender use of technical devices (s)	Women can use technical devices as well as men. (4)	–	–
Attitude towards marriage (s)	–	Do you think one should get married if one is living with a partner on a permanent basis? (3)	Marriage when living with partner permanent (4)
Family and labor-market outcomes			
Married (d)	Family status (4)	What is your marital status? (9)	What is your family status? (7)
Number of children (n)	Number of children	Do you have any children, and if so, how many?	Do you have or had children? If so, how much?
Labor-force participation (d)	Derived from: Currently employed? Currently unemployed? (2)	And now let's continue with employment and your occupation. Which of the categories on the card applies to you? (2)	Labor force status (11)
Employment (d)	Currently employed? (2)	And now let's continue with employment and your occupation. Which of the categories on the card applies to you? (2)	Labor force status (11)
Working hours (n)	How many hours per week do you actually work currently?	How many hours per week do you normally work in your main job, including overtime?	And how many hours do you generally work per week, including any overtime?
Earnings, log (n)	How high were your net earnings in your last month working? Please provide the sum after taxes and social insurance contributions. If you received extra compensation in your last month of working, such as vacation pay or back pay, please do not include this. Do, however, include overtime pay.	How high is your own net monthly income? By this I mean the amount remaining after deductions for tax and social security contributions.	What were your net earnings for the past month, after deductions for taxes and social insurance contributions, including overtime payments?

Notes: Continuation on next page

Outcome measures derived from the three datasets (continued)

	NEPS (1)	ALLBUS (2)	SOEP (3)
Educational outcomes			
Years of education (n)	Years of education = f(CASMIN)	Not counting the time you may have spent at a vocational school as part of your vocational training, how many years of schooling did you receive? If you went to university, please include the time you have spent there	Number of years of education
Abitur (d)	Which school-leaving certificate did you acquire? (8)	What general school leaving certificate do you have? (7)	What type of school-leaving certificate did you attain? (6)
Age of first employment (n)	Age at first employment (years)	–	How old were you when you first started working?

Notes: Translations of the original German questions from the official English codebook of the respective dataset. Scale of derived outcomes measure: s = standardized; n = number; d = dummy. For categorical variables, numbers in parentheses refer to the number of categories as presented in the respective dataset before recoding and merging. All gender and family role attitudinal outcomes are recoded such that an increase in the variable implies an increase in conservatism. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.3: Effects on religious outcomes: Non-parametric estimation

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Pre 15+	0.053	-0.005	0.056	0.069
	(0.128)	(0.901)	(0.043)	(0.035)
	[0.133]	[0.898]	[0.014]	[0.135]
Pre 10-14	0.055	0.032	0.038	0.033
	(0.253)	(0.343)	(0.202)	(0.216)
	[0.281]	[0.325]	[0.386]	[0.402]
Pre 5-9	0.007	-0.016	0.029	0.010
	(0.849)	(0.693)	(0.235)	(0.773)
	[0.881]	[0.782]	[0.334]	[0.812]
Post 1-5	-0.001	0.006	-0.004	-0.059
	(0.925)	(0.845)	(0.833)	(0.015)
	[0.917]	[0.833]	[0.836]	[0.078]
Post 6-10	-0.070	-0.044	-0.032	-0.035
	(0.085)	(0.016)	(0.272)	(0.211)
	[0.185]	[0.008]	[0.342]	[0.234]
Post 11-15	-0.125	-0.134	-0.092	-0.088
	(0.021)	(0.002)	(0.013)	(0.008)
	[0.076]	[0.003]	[0.004]	[0.026]
Post 16+	-0.168	-0.153	-0.178	-0.154
	(0.005)	(0.001)	(0.063)	(0.005)
	[0.068]	[0.007]	[0.113]	[0.015]
Observations	15,688	13,276	42,776	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.4: Effect on religiosity: Dummy coding of religiosity

	(Rather or very) religious		Very religious	
	Linear probability model	Probit model	Linear probability model	Probit model
	(1)	(2)	(3)	(4)
Reform	-0.029	-0.029	-0.022	-0.021
	(0.066)	(0.039)	(0.007)	(0.001)
	[0.124]	[0.073]	[0.064]	[0.092]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,688	15,688	15,688	15,688

Notes: Columns (1) and (3): OLS; columns (2) and (4): average marginal treatment effect of probit model. Dependent variable: columns (1) and (2): dummy equaling one if respondent is rather religious or very religious, zero otherwise; columns (3) and (4): dummy equaling one if respondent is very religious, zero otherwise. ALLBUS religiosity scale (from 1 to 10) re-scaled as very religious = 9-10 and rather religious = 6-8. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.5: Effect on religiosity: Controls for other school reforms

	Compulsory schooling	G8/G9	Philosophy	Sexual education	Political education	All other school reforms
	(1)	(2)	(3)	(4)	(5)	(6)
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.072	-0.073	-0.085	-0.071	-0.065	-0.067
	(0.040)	(0.017)	(0.014)	(0.048)	(0.041)	(0.067)
	[0.058]	[0.058]	[0.073]	[0.236]	[0.140]	[0.135]
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,198	15,198	15,185	15,198	15,198	15,198

Notes: Dependent variable: Standardized religiosity (see Table A2.2 for details). Regressions include additional controls for school reforms as enacted in the state and year of a respondent's primary school entry, as indicated in the column header: (1) years of compulsory schooling (between 8 and 10); (2) dummy equaling one if duration of Gymnasium is 8 years, zero otherwise; (3) dummy equaling one if philosophy is taught in school (above and beyond the school subject "ethics" evaluated in this paper), zero otherwise; (4) dummy equaling one if sexual education is taught in school, zero otherwise; (5) dummy variable equaling one if political education is taught in school, zero otherwise; (6) all five school reforms together. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.6: Effect on personal prayer: Controls for other school reforms

	Compulsory schooling	G8/G9	Philosophy	Sexual education	Political education	All other school reforms
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.063	-0.054	-0.061	-0.054	-0.040	-0.068
	(0.098)	(0.059)	(0.077)	(0.060)	(0.098)	(0.063)
	[0.069]	[0.072]	[0.134]	[0.062]	[0.084]	[0.058]
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,929	12,929	12,915	12,929	12,929	12,929

Notes: Dependent variable: Standardized personal prayer (see Table A2.2 for details). Regressions include additional controls for school reforms as enacted in the state and year of a respondent's primary school entry, as indicated in the column header: (1) years of compulsory schooling (between 8 and 10); (2) dummy equaling one if duration of Gymnasium is 8 years, zero otherwise; (3) dummy equaling one if philosophy is taught in school (above and beyond the school subject "ethics" evaluated in this paper), zero otherwise; (4) dummy equaling one if sexual education is taught in school, zero otherwise; (5) dummy variable equaling one if political education is taught in school, zero otherwise; (6) all five school reforms together. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.7: Effect on church-going: Controls for other school reforms

	Compulsory schooling	G8/G9	Philosophy	Sexual education	Political education	All other school reforms
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.059	-0.054	-0.076	-0.049	-0.042	-0.049
	(0.055)	(0.041)	(0.018)	(0.094)	(0.103)	(0.078)
	[0.044]	[0.042]	[0.062]	[0.144]	[0.134]	[0.102]
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,559	41,559	41,457	41,559	41,559	41,559

Notes: Dependent variable: Standardized church-going (see Table A2.2 for details). Regressions include additional controls for school reforms as enacted in the state and year of a respondent's primary school entry, as indicated in the column header: (1) years of compulsory schooling (between 8 and 10); (2) dummy equaling one if duration of Gymnasium is 8 years, zero otherwise; (3) dummy equaling one if philosophy is taught in school (above and beyond the school subject "ethics" evaluated in this paper), zero otherwise; (4) dummy equaling one if sexual education is taught in school, zero otherwise; (5) dummy variable equaling one if political education is taught in school, zero otherwise; (6) all five school reforms together. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.8: Effect on religious affiliation: Controls for other school reforms

	Compulsory schooling	G8/G9	Philosophy	Sexual education	Political education	All other school reforms
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.081	-0.087	-0.096	-0.088	-0.080	-0.075
	(0.012)	(0.004)	(0.007)	(0.003)	(0.002)	(0.009)
	[0.025]	[0.032]	[0.053]	[0.039]	[0.021]	[0.252]
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44,571	44,571	44,465	44,571	44,571	44,571

Notes: Dependent variable: Standardized religious affiliation (see Table A2.2 for details). Regressions include additional controls for school reforms as enacted in the state and year of a respondent's primary school entry, as indicated in the column header: (1) years of compulsory schooling (between 8 and 10); (2) dummy equaling one if duration of Gymnasium is 8 years, zero otherwise; (3) dummy equaling one if philosophy is taught in school (above and beyond the school subject "ethics" evaluated in this paper), zero otherwise; (4) dummy equaling one if sexual education is taught in school, zero otherwise; (5) dummy variable equaling one if political education is taught in school, zero otherwise; (6) all five school reforms together. Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.9: Effects on religious outcomes: Dosage treatment

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.092	-0.047	-0.074	-0.097
	(0.016)	(0.146)	(0.032)	(0.012)
	[0.053]	[0.156]	[0.042]	[0.010]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,688	13,276	42,776	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.10: Effects on religious outcomes: Reform timing coded at secondary school entry

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.071	-0.041	-0.058	-0.070
	(0.040)	(0.154)	(0.035)	(0.035)
	[0.121]	[0.128]	[0.050]	[0.133]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,688	13,276	42,776	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.11: Effects on religious outcomes: Excluding early reforming states

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.090	-0.023	-0.121	-0.087
	(0.014)	(0.608)	(0.062)	(0.037)
	[0.086]	[0.538]	[0.365]	[0.342]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	8,320	7,046	23,039	24,245

Notes: Observations from Bavaria, Hesse, Lower Saxony, and Rhineland-Palatinate are excluded from the sample. Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.12: Effects on religious outcomes: Only NEPS data

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.090	-0.083	Not covered	-0.071
	(0.002)	(0.023)	in NEPS	(0.161)
	[0.012]	[0.034]		[0.321]
State fixed effects	Yes	Yes		Yes
Birth-year fixed effects	Yes	Yes		Yes
Controls	Yes	Yes		Yes
Observations	9,232	7,963		9,237

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data source: National Education Panel Study (NEPS) Cohort 6.

Table A2.13: Effects on religious outcomes: Only ALLBUS data

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.044	0.018	-0.062	-0.111
	(0.326)	(0.635)	(0.077)	(0.001)
	[0.438]	[0.677]	[0.175]	[0.026]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	6,456	5,313	15,714	15,860

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data source: German General Social Survey (ALLBUS) Cumulation 1980-2016.

Table A2.14: Effects on religious outcomes: Only SOEP data

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	Not covered	Not covered	-0.055	-0.066
	in SOEP	in SOEP	(0.065)	(0.058)
			[0.042]	[0.139]
State fixed effects			Yes	Yes
Birth-year fixed effects			Yes	Yes
Controls			Yes	Yes
Observations			27,062	20,828

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data source: German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.15: Effects on religious outcomes: Callaway and Sant'Anna estimator

	Religiosity		Prayer		Church-going		Affiliation	
	Without controls	With controls	Without controls	With controls	Without controls	With controls	Without controls	With controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform	-0.086	-0.114	-0.109	-0.131	-0.097	-0.105	-0.078	-0.098
	(0.061)	(0.042)	(0.136)	(0.052)	(0.112)	(0.095)	(0.000)	(0.000)
Observations	15,066	15,063	12,821	12,821	41,232	41,219	44,187	44,187

Notes: Simple average treatment effects based on Callaway and Sant'Anna (2021), with not yet treated units and never-treated units as controls. Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls (if included, as indicated in the column header): gender, migration status, mother's education, father's education, and survey fixed effects. Estimator: DR IPW estimator. Inference: p-values with clustering at the state level. Implemented using Stata package csdid, version 1.6 (Rios-Avila et al., 2021). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

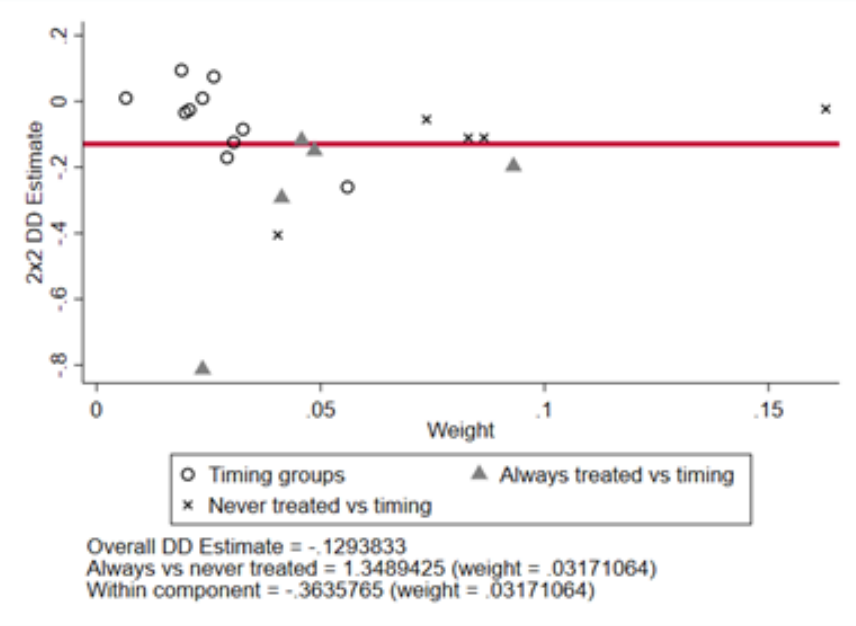
Table A2.16: Effects on religious outcomes: Excluding Berlin and Bremen

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.073	-0.050	-0.052	-0.087
	(0.028)	(0.084)	(0.059)	(0.007)
	[0.095]	[0.097]	[0.058]	[0.051]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,066	12,821	41,232	44,193

Notes: Observations from Berlin and Bremen are excluded from the sample. Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.17: Goodman-Bacon decomposition of effect on religiosity

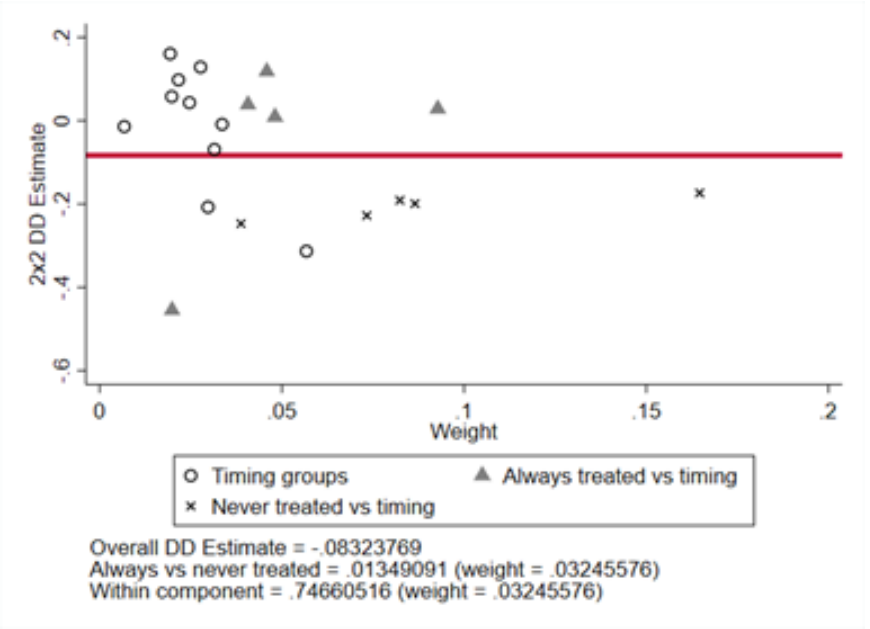
	Beta (1)	Total weight (2)
Overall	-0.129	-
Timing groups	-0.087	0.263
Always vs. timing	-0.247	0.252
Never vs. timing	-0.095	0.446
Always vs. never	1.349	0.006
Within	-0.364	0.032



Notes: Decomposition of difference-in-differences estimator with variation in treatment timing based on Goodman-Bacon (2021). The figure shows a scatterplot of all two-group/two-period difference-in-difference estimates and their associated weights in the two-way fixed effects model. Depicted types differ by control group: (1) timing groups, or groups whose treatment at different times serves as each other's control groups; (2) always treated, where a group treated prior to the start of the analysis serves as the control group; and (3) never treated, where a group which never receives the treatment serves as the control group. Also shown are the component due to variation in controls across always treated and never treated groups, as well as the "within" residual component. Data are collapsed to means of state-cohort cells. Observations with birth year before 1949 or after 1991 are dropped to create a balanced panel. Dependent variable: religiosity (standardized). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.18: Goodman-Bacon decomposition of effect on personal prayer

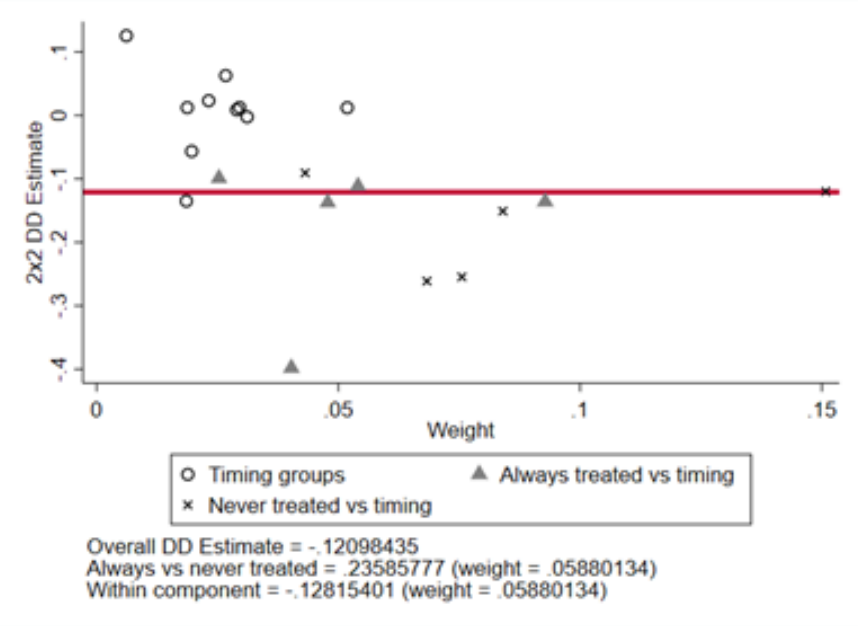
	Beta (1)	Total weight (2)
Overall	-0.083	-
Timing groups	-0.056	0.271
Always vs. timing	0.004	0.247
Never vs. timing	-0.196	0.445
Always vs. never	0.013	0.005
Within	0.747	0.032



Notes: Decomposition of difference-in-differences estimator with variation in treatment timing based on Goodman-Bacon (2021). The figure shows a scatterplot of all two-group/two-period difference-in-difference estimates and their associated weights in the two-way fixed effects model. Depicted types differ by control group: (1) timing groups, or groups whose treatment at different times serves as each other's control groups; (2) always treated, where a group treated prior to the start of the analysis serves as the control group; and (3) never treated, where a group which never receives the treatment serves as the control group. Also shown are the component due to variation in controls across always treated and never treated groups, as well as the "within" residual component. Data are collapsed to means of state-cohort cells. Observations with birth year before 1949 or after 1991 are dropped to create a balanced panel. Dependent variable: prayer (standardized). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.19: Goodman-Bacon decomposition of effect on church-going

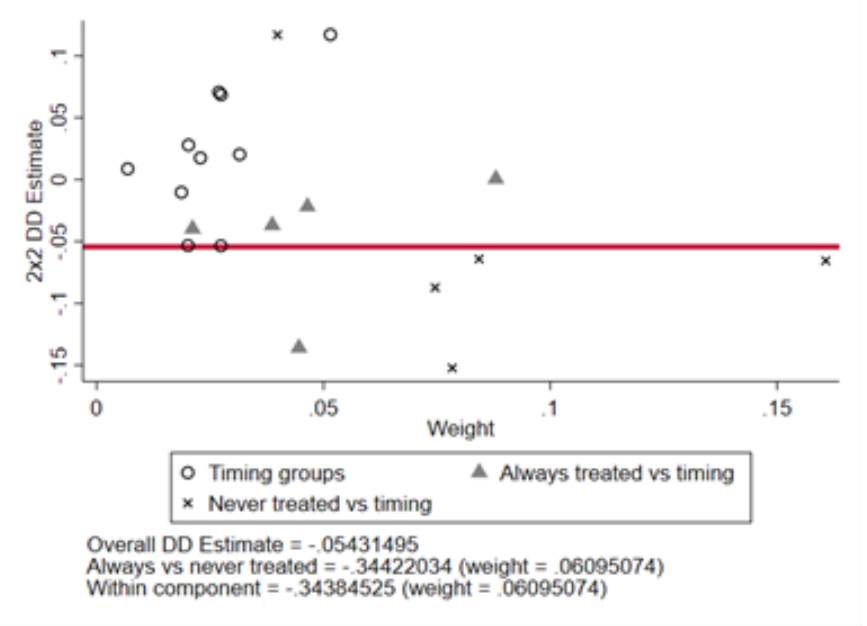
	Beta (1)	Total weight (2)
Overall	-0.121	-
Timing groups	0.003	0.254
Always vs. timing	-0.168	0.260
Never vs. timing	-0.169	0.421
Always vs. never	0.236	0.006
Within	-0.128	0.059



Notes: Decomposition of difference-in-differences estimator with variation in treatment timing based on Goodman-Bacon (2021). The figure shows a scatterplot of all two-group/two-period difference-in-difference estimates and their associated weights in the two-way fixed effects model. Depicted types differ by control group: (1) timing groups, or groups whose treatment at different times serves as each other's control groups; (2) always treated, where a group treated prior to the start of the analysis serves as the control group; and (3) never treated, where a group which never receives the treatment serves as the control group. Also shown are the component due to variation in controls across always treated and never treated groups, as well as the "within" residual component. Data are collapsed to means of state-cohort cells. Observations with birth year before 1949 or after 1991 are dropped to create a balanced panel. Dependent variable: church-going (standardized). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.20: Goodman-Bacon decomposition of effect on religious affiliation

	Beta (1)	Total weight (2)
Overall	-0.054	-
Timing groups	0.035	0.253
Always vs. timing	-0.039	0.239
Never vs. timing	-0.068	0.438
Always vs. never	-0.344	0.010
Within	-0.344	0.060



Notes: Decomposition of difference-in-differences estimator with variation in treatment timing based on Goodman-Bacon (2021). The figure shows a scatterplot of all two-group/two-period difference-in-difference estimates and their associated weights in the two-way fixed effects model. Depicted types differ by control group: (1) timing groups, or groups whose treatment at different times serves as each other's control groups; (2) always treated, where a group treated prior to the start of the analysis serves as the control group; and (3) never treated, where a group which never receives the treatment serves as the control group. Also shown are the component due to variation in controls across always treated and never treated groups, as well as the "within" residual component. Data are collapsed to means of state-cohort cells. Observations with birth year before 1949 or after 1991 are dropped to create a balanced panel. Dependent variable: affiliation (standardized). Controls: gender, migration status, mother's education, father's education, survey and survey-year fixed effects. Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

Table A2.21: Effects on religious outcomes: No control variables

	Religiosity	Prayer	Church-going	Affiliation
	(1)	(2)	(3)	(4)
Reform	-0.073	-0.038	-0.068	-0.079
	(0.011)	(0.181)	(0.024)	(0.015)
	[0.032]	[0.217]	[0.026]	[0.087]
State fixed effects	Yes	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes	Yes
Survey-year fixed effects	Yes	Yes	Yes	Yes
Observations	15,688	13,276	42,776	45,925

Notes: Dependent variables indicated in column headers. All dependent variables are standardized (see Table A2.2 for details). Inference: p-values with clustering at the state level; parentheses: standard clustering at state level; brackets: wild cluster bootstrap by Roodman et al. (2019). Data sources: National Education Panel Study (NEPS) Cohort 6; German General Social Survey (ALLBUS) Cumulation 1980-2016; German Socio-Economic Panel (SOEP) Core 1984-2017 (v.34).

3 Evolution vs. Creationism in the Classroom: The Lasting Effects of Science Education*

3.1 Introduction

Anti-scientific attitudes can impose substantial costs on public health, the environment, and the economy. Misinformation about the danger of Covid-19 and a lack of trust in scientists and has undermined compliance with social distancing measures and vaccination recommendations, prolonging the pandemic (Bursztyn et al., 2020; Algan et al., 2021; Brzezinski et al., 2021; Jin et al., 2021). Climate change denial has reduced the support for policies cutting greenhouse gas emissions, contributing to its environmental and economic damage (Akter et al., 2012; Linden et al., 2015). The rejection of evolution theory has been used to justify white supremacy and racism in the US (Marks, 2012), and has contributed to anti-scientific agricultural policies and resulting food shortages in the Soviet Union (Graham, 2016).[1] While the societal costs of anti-scientific attitudes are well understood, evidence on its determinants is surprisingly scant despite its relevance for effective policy responses.

This paper isolates the content of high school science education as one determinant of anti-scientific attitudes that is directly subject to the policy maker.[2] To study whether the content of science education has lasting effects on individuals beyond attitudinal outcomes, the paper also analyzes how it affects scientific knowledge and life decisions. Specifically, I estimate the causal effect of students' exposure to the teaching of evolution theory in science education on (i) their knowledge about evolution at the end of schooling, (ii) their attitudes on evolution in adulthood, and (iii) the probability that they work in life sciences.

The focus of this paper is on evolution theory because of its fundamental role in science, and its controversy in the population and the education system. Evolution can scientifically explain the existence of all species including our own. The American Association for the Advancement of Science (2021) states that "the foundation of all life sciences is biological evolution". 98 percent of its members express support for the statement that humans have

* This chapter is based on the job market paper 'Evolution vs. Creationism in the Classroom: The Lasting Effects of Science Education', mimeo.

1 The pseudoscientific theories of Trofim Lysenko, then-president of the Academy of Agricultural Sciences of the USSR and leading agricultural advisor to Joseph Stalin, have been made responsible for prolonging Soviet foot shortages in the 1930s ("Lysenkoism") (Joravsky, 1962).

2 In general, attitudes are shaped by a multitude of factors many of which are rather shielded in the private domain. A large literature on the formation of attitudes and beliefs has emphasized the impact of inter-generational transmission in families (Bisin and Verdier, 2001; Guiso et al., 2008; Tabellini, 2008). Other determinants include peers and social networks (Sacerdote, 2001; Bailey et al., 2020), the media (Martin et al., 2017), political systems (Alesina and Fuchs-Schündeln, 2007), and macroeconomic conditions (Giuliano and Spilimbergo, 2014).

evolved over time (Pew Research Center, 2015). In contrast, evolution is a highly charged topic in the US population with only 65 percent agreeing that humans have evolved over time. This controversy has been reflected in heated debates and legal battles from before World War I to the present day on whether evolution is supposed to be taught in schools.[3] Teachers and school districts have been convicted for not following the educations standards' stance on evolution. Even today, there is substantial variation across US states and years in the way how evolution is covered in education standards.

To isolate exogenous variation in students' exposure to the teaching of evolution, this paper exploits staggered state-level reforms of the coverage of evolution in US State Science Education Standards (Science Standards). In the study period from 2000 until 2009, 22 states expanded the coverage of evolution in their education standard, and 15 states reduced it. I argue that the political and institutional processes leading to these reforms, in particular the predetermined timing of gubernatorial elections and the tenure of members of State Boards of Education, create idiosyncrasies in the determination of the precise reform years. This setting allows for the estimation of causal effects in two-way fixed effects models with state and cohort fixed effects, overcoming the identification problem that the content of science education is generally correlated with scientific, religious and political attitudes of the students' environment which independently affect student outcomes.

Beyond the theoretical argument that the reform timing is determined by institutional idiosyncrasies, my empirical setup takes explicitly care of a range of endogeneity concerns by comparing adjacent cohorts around sharp reforms of the State Standards. Specifically, the performed two-way fixed effects estimations can rule out as confounding factors (i) state-specific differences (such as education levels), (ii) cohort-specific differences (such as national changes in attitudes across time), (iii) time-varying state-specific shocks that affect adjacent cohorts similarly (such as natural disasters or state-level political or religious shocks that do not differentially affect children of different cohorts), and (iv) time-varying state-specific shocks that affect adjacent cohorts differentially, but smoothly (such as state-specific trends in science skepticism), in a robustness test that includes state-specific time trends. To conduct the set of analyses, I link state-level data on the evolution coverage in Science Standards with three individual-level datasets.

First, this paper shows that the evolution coverage in Science Standards affects what students learn about evolution in school. Specifically, I use the National Assessment for Educational Progress (NAEP) to demonstrate that students being exposed to a more comprehensive evolution coverage in high school are more likely to answer knowledge questions on evolution correctly by the end of high school. This finding exemplifies how the content of education standards can foster scientific knowledge, an outcome of direct economic importance given its

[3] For example, the New York Times published a report on recent controversies with the headline "Questioning Evolution: The Push to Change Science Class" (Haberman, 2017).

effects on earnings and economic growth in the long run (Lucas, 1988; Barro, 2001; Hanushek and Woessmann, 2008, 2012).

Second, this paper demonstrates that the evaluated reforms have lasting effects on attitudes. To that end, I make use of the General Social Survey (GSS) to show that evolution teaching affects the probability to approve of the concept of evolution in adulthood. Being exposed in high school to a comprehensive evolution coverage in the education standard as opposed to no evolution coverage increases evolution approval in adulthood by 57 percent of the sample mean. This analysis underscores that reform effects persist long after students have left high school. This result exemplifies how science education can promote scientific attitudes, which can be directly relevant for improving public health, the environment, and the economy (Brzezinski et al., 2021; Martinez-Bravo and Stegmann, 2021).

Third, this paper shows that the evaluated reforms affect high-stakes choices, namely occupational choice. I hypothesize that learning about evolution, the fundamental theory of life sciences, affects the probability to work in life sciences in adulthood. Using the American Community Survey (ACS), I demonstrate that being exposed in high school to a comprehensive evolution coverage in the education standard as opposed to no evolution coverage increases the probability to work in life sciences in adulthood by 23 percent of the sample mean. This effect is mostly coming from the subgroup of biology, the subject in which evolution is typically being taught. This finding exemplifies how science education can attract future STEM workers, which does not only raise wages at the individual level (Hastings et al., 2013; Kirkeboen et al., 2016; Deming and Noray, 2020), but also has wider economic consequences through fostering innovation, technological change, labor productivity and economic growth (Griliches, 1992; Jones, 1995; Kerr and Lincoln, 2010; Peri et al., 2015).

A particularly useful feature of focusing on one topic such as evolution is the possibility to construct fine-grained placebo tests. Testing whether reforms affect non-evolution outcomes constitutes falsification tests. Specifically, I show null effects for (i) non-evolution scientific knowledge by the end of high school, (ii) non-evolution scientific, religious, and political attitudes in adulthood, and (iii) the probabilities to work in non-scientific occupational fields. These results provide empirical support for the interpretation that it is indeed institutional idiosyncrasies which determine the exact reform timing and not scientific, religious, and political trends or shocks. I further demonstrate that the reform effect on evolution knowledge is specific to students in public schools, while there is no effect for a placebo sample of private school students for whom Science Standards have never been binding. I also present event-study graphs which show no pre-trends. They also demonstrate that reform effects occur soon after reform adoption and are more pronounced for the group of states which reduce their evolution teachings relative to those which expand their teachings. Another robustness check replicates the main results on a subsample using only states with closely elected governors ruling at the time of the reform. In addition, the results are immune to potential biases in staggered two-way fixed effects designs from time-varying treatment effects (Callaway and Sant'Anna, 2021).

This paper contributes to the literature on the political economy of schooling (Lott, 1999; Cantoni and Yuchtman, 2013). I provide the first quasi-experimental evidence that attitudinal changes induced by reforms of the content of education translate into high-stakes choices of individuals. Cantoni et al. (2017) exploit a Chinese textbook reform to show that the content of education affects students' political and economic attitudes. Other seminal papers study the effects of the content of education on cultural identity (Clots-Figueras and Masella, 2013), civic values (Bandiera et al., 2019), and religiosity (Bazzi et al., 2020; Arold et al., 2022). While these papers show effects on attitudes, I go beyond that and demonstrate that high-stakes occupational choice is also affected.

At the same time, this finding enhances our understanding of how to increase the share of STEM graduates, which is a widely shared policy goal in many societies.[4] We know that occupational sorting is influenced by demand side factors such as expected earnings and non-pecuniary job benefits (Wiswall and Zafar, 2018; Arcidiacono et al., 2020), perceived ability (Stinebrickner and Stinebrickner, 2014; Arcidiacono et al., 2016a), and heterogeneous tastes (Wiswall and Zafar, 2015). Supply side factors such as grading policies (Butcher et al., 2014), admissions systems (Bordon and Fu, 2015), affirmative action policies (Arcidiacono et al., 2016b), or the provision of role models (Jackson et al., 2020) can also play a role (for an overview, see also (Altonji et al., 2016)). We demonstrate that the content of science education in high school can be an effective policy tool to attract STEM graduates.

This paper also speaks to the emerging literature on the determinants of religiosity (Iannaccone, 1998; Iyer, 2016; McCleary and Barro, 2019). Finding null effects on religious outcomes demonstrates that neither believing in nor belonging to religions is reduced by expanding the scientific content of science education. This is true despite the fact that being raised as Evangelical is the largest negative predictor of evolution approval in this study. While a number of studies have found a positive relationship between education and religiosity (McCleary and Barro, 2006a; Glaeser and Sacerdote, 2008; Meyersson, 2014), other research suggests that education can decrease religiosity (Hungerman, 2014; Becker et al., 2017). In the specific setting of evolution teaching in the US, religiosity is not crowded out.

Lastly, this paper contributes to the literature on the effects of content of education on students' knowledge. While the benefits of increasing topic-specific instruction time (Cortes and Goodman, 2014), minimum high school course requirements (Goodman, 2019), and advanced placement courses (Conger et al., 2021) are well understood, this paper can show that the content of education standards affects the knowledge of students in the intended direction.[5] Even more, the effects of the content of education standards last into adulthood.

[4] In the US, increasing the number of STEM graduates is a central policy goal of the Federal Government's strategic plan for STEM education 2018-2023 (National Science and Technology Council, 2018). Similarly, the EU aims to increase the number of STEM graduates as one of its twelve policy goals of the European Skills Agenda 2020-2025 (European Commission, 2020).

[5] Arold and Shakeel (2021) show that the adoption of centralized education standards in the US in math and ELA had *unintended* effects on students' science achievement.

The paper proceeds as follows. Section 3.2 outlines the historical and institutional background of the teaching of evolution. Section 3.3 provides information on the data measuring the coverage of evolution in Science Standards and the microeconometric datasets. Section 3.4 describes the identification strategy. Section 3.5 presents the results. Section 3.6 discusses robustness tests. Section 3.7 concludes.

3.2 Institutional Background

3.2.1 The Battle for Teaching Evolution in US Public Schools

The teaching of evolution in public schools has been a contested issue for at least a century in the US. Although the scientific community reached a consensus on the validity of evolution relatively soon after Charles Darwin's publication of "On the Origin of Species" (Darwin, 1859),[6] the public did not share the consensus. This was and still is reflected in the educational system. For the decades before World War I, Beale (1941) describes that teachers wanting to teach evolution in an average American school had difficulties to do so. Only one quarter of the biology textbooks published between 1900 and 1919 contained information about evolution (Skoog, 2005). No book covered human evolution. In the 1920s, about one third of biology textbooks covered human evolution, documenting an early phase of a gradual and non-linear development throughout the 20^{th} century towards more evolution coverage in US high school biology textbooks.

However, the 1920s also marked the start for a series of legal disputes about the teaching of evolution in US schools throughout the 20^{th} century. At least 20 states considered bills to ban the coverage of evolution in public schools in the 1920s (Numbers, 1982). Among other states, such a bill became law in Tennessee, known as the Butler Act, which resulted in the famous Scopes trial in 1925. John T. Scopes, a biology teacher from Tennessee, was convicted in Tennessee v. Scopes for having taught evolution in the classroom. Although the Tennessee Supreme Court overturned the decision on a technicality, it decided that the law banning evolution from schools was not unconstitutional (Larson, 1999).

In the second half of the 20^{th} century, legislative and adjudicative decisions became more favorable towards the coverage of evolution in public schools (Moore et al., 2003b). In 1967, the Butler's Act was repealed by the Tennessee legislature. One year later, the Supreme Court of the US ruled that a law banning the teaching of evolution in schools in Arkansas was unconstitutional in Epperson v. Arkansas. As a reaction, creationists lobbied for laws requiring that equal time must be spent on teaching evolution and creation. In 1987, this was

[6] Thomas Henry Huxley (1880, p.1) stated that "there is no field of biological inquiry in which the influence of the 'Origin of Species' is not traceable [...] and the general doctrine of evolution [...] may conduct its conquest of the whole realm of Nature". Ernst Mayr (1991, p.25), a leading evolutionary biologists of the 20^{th} century, wrote that "within fifteen years of the publication of the Origin hardly a qualified biologist was left who had not become an evolutionist".

ruled unconstitutional by the US Supreme Court in Edwards v. Aguillard. In sum, the legal decisions of the 20^{th} century have paved the way for evolution to be taught in public schools. In the 21^{st} century, creationism and intelligent design are not permitted to be taught in US public schools anymore. Still, there remains substantial variation in evolution teaching across states and years, as the subsequent analysis of the evolution coverage in Science Standards demonstrates.

3.2.2 US State Science Standards

US State Science Standards serve as state-wide school curriculum frameworks in science. Historically, the content of US education has been determined at the local level. However, concerns about achievement declines of US students in the 1960s and 1970s and resulting economic costs (Hanushek, 1986; Bishop, 1989) gave rise for calls to establish rigorous and comparable education standards. In 1983, the report "A Nation at Risk" (National Commission on Excellence in Education, 1983) proposed the introduction of centralized education standards.[7] Several organizations have proposed guidelines for centralized educational standards for the different school subjects. Regarding science, the American Association for the Advancement of Science developed the Science Standard guidelines "Science for All Americans" (1990) and "Benchmarks for Science Literacy" (1994), and the National Research Council published the "National Science Education Standards" (1996). By 2000, all states except for Iowa had adopted Science Standards (Lerner, 2000a).

Science Standards define the scientific knowledge and skills students are supposed to master in a given grade in public schools. The scientific teaching a student is ultimately exposed to in class does not solely depend on the Science Standard of her state, but also on local school curricula, the selection of textbooks, the knowledge, ability and ideology of teachers, testing formats, and other factors. However, Science Standards form the basis of many of these factors and thereby, indirectly, affect the science teaching in schools. For instance, they affect how local curricula and lesson plans of teachers are written (Lerner, 2000b). Furthermore, science textbooks are arranged to match the content laid out in Science Standards, particularly reflecting the standards from larger states. In addition, statewide standardized exams often directly test the content set out in the Science Standards. Lerner (2000b, p.ix) summarizes that "the knowledge and skills set forth in state standards are supposed to form the core of "standard based" education reform. They are meant to serve as the frame to which everything else is attached, the desired outcome that drives countless other decisions about how best to attain it." Overall, 88 percent of a nationally representative sample of US public high school

[7] Theoretically, centralized education standards can be more rigorous as they overcome a free-riding problem induced by the mobility of high school graduates across school districts and their pooling in the local labor markets (Costrell, 1994, 1997). At the same time, centralization can also reduce the incentive to develop rigorous and innovative education standards by abolishing competition between school districts (Tiebout, 1956; Oates, 1999).

biology teachers state that they focus heavily on what students need to know to meet Science Standards when teaching evolution (Figure A3.1).

3.2.3 Reform Examples from Florida and Texas

Reforms of the evolution coverage in Science Standards form the basis of the two-way fixed effects design performed in this paper. The following two reform examples illustrate how such reforms come into existence. While Florida expanded the evolution coverage in 2008, Texas reduced it in 2009. The Science Standard in power in Florida before 2008 did not mention the word "evolution", and its discussion of evolutionary processes (under a different wording) were minimal.[8] In February 2008, the Florida Board of Education voted 4:3 in favor of a new Science Standard that included evolution comprehensively. This close majority emerged after years of debating and drafting the Standard. In fact, the Standard was re-drafted yet another time just hours before the final vote. Replacing the term "evolution" by "the scientific theory of evolution" secured the majority ultimately. The new Standard captured biological, geological, cosmological and even human evolution comprehensively (Mead and Mates, 2009).

In contrast to Florida, Texas reduced the evolution coverage in 2009. The evolution coverage in the Science Standard in place in 2000 was described as "brief but satisfactory" (Lerner, 2000b, p.15). It contained all areas of evolution except for human evolution. In 2003, Don McLeroy, the then-chairman of the Texas Board of Education, pushed for a much more limited evolution coverage. He stated that he personally does not believe in Darwin's evolution theory and in the earth being older than a couple of thousand years which was in part reflected in the Science Standard proposal. In 2003, his reform proposal found no majority in the Board of Education, and years of debate followed. In 2009, he proposed another Science Standard which required that "strengths and weaknesses" of evolution should be taught. This was regarded by some as an attempt to open up room for the teaching of creationism at the teachers' discretion, without mentioning creationism explicitly in the Science Standard. It was voted down 8-7. A second version required students to study the "sufficiency or insufficiency" of key principles of evolution. It was also voted down 8-7. A third attempt which contained creationist jargon more subtly was ultimately approved by 13-2 votes. This new Science Standard left out some areas of the teaching of evolution and added "pieces of creationist jargon" (Mead and Mates, 2009, p.366). For example, the phrase that "the estimated age of the universe was 14 billion years" was removed. Notably, the reforms in Florida and Texas did not follow a partisan change as all governors in Florida and Texas in the 21^{st} century have been Republican. Both reform examples shed light on the political process behind such reforms, and show that they are not simply a consequence of a change of government.

[8] Lerner (2000b, p.14) describes the Science Standard as "Extensive standards that skim lightly over biological and geological evolution without ever mentioning the word. Not satisfactory."

3.2.4 The Adoption Process of Reforms of Science Standards

Understanding the political process leading to reforms of the evolution coverage in Science Standards is of particular interest for assessing whether they create exogenous variation in students' exposure to the teaching of evolution. In this section, I argue that the fact that such reforms happen *at some point* is not as-good-as-random, but that the *specific timing* of such reforms is as-good-as-random given substantial institutional idiosyncrasies.

Reforms of Science Standards are decided on by majority vote of the members of the State Boards of Education. The selection process of the members of the State Boards of Education differs across states. In some states, the members are appointed by the governor, sometimes with the advice and consent of the senate (for example in California and Florida). In other states, the State Boards of Education members are elected by the public, typically in a staggered election across districts (for example in the District of Columbia and Texas). Some states combine the two selection mechanisms by appointing some members and electing others (for example Louisiana and Ohio). Student representatives or external experts are also appointed or elected in some states (for example in Alaska and Massachusetts).

Before the members of the State Board of Education vote on a reform of Science Standards, the standards are typically drafted by advisory committees. The composition of these advisory committees depends, again, on the state. In general, advisory committees consist of a panel of teachers and other stakeholders including, at least occasionally, scientists. In addition to the input of the advisory board, most State Boards of Education hold hearings or testimonies of stakeholders such as parents, scientists, religious representatives, among others. At this point, it typically becomes visible which interest groups lobby in favor or against the proposed reform. For example, the National Center for Science Education has lobbied for a more comprehensive coverage of evolution in multiple cases, while the Discovery Institute has spoken out against it. The probability of anti-evolution organizations being active in a state depends, among others, on the state's conservative Protestant adherence rate (Johnson et al., 2016). After the period of public comment, the State Boards of Education has the final vote.

On the one hand, the political process described above implies that these reforms happen at some point in a given state is not random. Instead, they reflect changing political views, either expressed by the election of a governor who subsequently appoints members of the State Boards of Education, or by direct election of the members of the State Boards of Education.

On the other hand, the exact reform year in a given state can be regarded as-good-as random. If the approval towards evolution or science in general changes in the population in a certain year, it will take a somewhat arbitrary number of years until this results in a reform of Science Standards because of institutional idiosyncrasies. In states where members of the Board of Education are appointed by the governor, the year of a reform crucially depends on the election year of governors, which is determined by the legislation period lasting four years in general. In states where members of the Board of Education are directly elected, the reform

year depends on the elections, which typically take place in different districts in different years in a staggered manner. Further state-specific idiosyncrasies are induced by the fact that the tenure of members of the Boards of Education differs across states, which can last up to nine years like in West Virginia. Even after a new majority in the Board of Education is in power, the drafting, hearing, and voting on new standards causes further delay, as this can take months or years. In some cases, there are also spillovers in the sense that Science Standards reforms of one state affect the teaching in other states. This occurs, for example, because textbooks used in smaller states may follow Science Standards reforms of larger states. In sum, the number of years between a scientific, religious, or political shock and a reform of the evolution coverage in Science Standards can be very large. However, it can also be small if election dates and the tenure expiration of the marginal board member happen shortly after a given shock. Hence, the precise timing of such reforms is arguably exogenous. In the empirical analysis, placebo tests showing null effects on non-evolution scientific, religious and political outcomes test this narrative empirically. The same is true for regressions conditioning on the party of the governor.

3.2.5 The Implementation of Reforms of Science Standards

After new Science Standards are adopted, their implementation in the classroom tends to be rather swift. In general, widely publicized lawsuits convicting school districts for not implementing the teaching of evolution as outlined in Science Standards contribute to a fast implementation of such reforms.[9] In Florida in 2008, for example, school districts were supposed to adjust their lessons by including evolution comprehensively as outlined in the newly adopted Science Standard within one year. Furthermore, evolution was required to become part of standardized testing in Florida from 2012 onwards. In the 2009 Texas reform, the evolution coverage of the new Science Standard had to be in textbooks from 2011 onwards.

3.3 Data

3.3.1 Coding of Reforms of Science Standards

To measure the coverage of evolution in a Science Standard, I make use of the "evolution score" provided by Lerner (2000a) and Mead and Mates (2009). The evolution score is a composite index based on an evaluation of whether the word "evolution" appears in a Science Standard,

[9] For example, a lawsuit that received national attention was Kitzmiller v. Dover Area School District in 2005. The Dover Area School District had required biology teachers to teach intelligent design (a form of creationism attributing the creation of the world to an intelligent designer) as an alternative to evolution. This requirement contradicted the content of the Science Standard in power at the time, and was ruled unconstitutional in Kitzmiller v. Dover Area School District. Specifically, the verdict barred the Dover Area School District from requiring teachers to "denigrate or disparage the scientific theory of evolution, and from requiring teachers to refer to a religious, alternative theory known as intelligent design." (Kitzmiller v. Dover Area School District, 400 F. Supp. 2d 707 (M.D. Pa. 2005)).

of the respective coverages of biological, human, geological, and cosmological evolution, and of the connection of the different aspects of evolution. In addition, the absence of creationist jargon and creationist disclaimers in textbooks is taken into account. The evolution score is defined between 0 and 1, with 0.01 increments. An evolution score of 0 indicates no or a non-scientific/creationist coverage of evolution, and a score of 1 a very comprehensive coverage of evolution. Notably, the creationist jargon in all Standards evaluated in this paper is never openly religious, which would be unconstitutional. However, there is large variation in the emphasis of (alleged) weaknesses and critique of evolution theory, opening or closing space for teachers who wish to teach creationist content.[10]

The evolution score is available for all states for the years 2000 and 2009, provided by Lerner (2000a) and Mead and Mates (2009), respectively. They also provide information on the year of reform of the evolution score for each state between 2000 and 2009 (if there was any reform). If more than one reform happened between 2000 and 2009 in a given state, there is information on the last reform.[11] The evolution score serves as treatment variable in all analyses presented in this paper. When merging it with individual-level datasets, each individual is defined to be exposed to the evolution score from 2000 if she entered high school before the reform year in her state, and to be exposed to the evolution score from 2009 if she entered high school in the year of the reform in her state or later. The high school entry year is the relevant year, as most of the teaching on evolution takes place at the beginning of high school.[12]

To illustrate the identifying variation, Figure 3.1 depicts the state-level evolution score *difference* between 2000 and 2009.[13] The evolution score increased in 22 states (implying a positive evolution score difference) and decreased in 15 states (implying a negative evolution score difference) between 2000 and 2009. In the remaining 13 states, it remained unchanged. The states with the largest evolution score increases are Kansas, Mississippi, and Florida. The largest evolution score decreases are found in Connecticut, Louisiana, and Texas. By

[10] In 2000, Kansas received an out-of-range score of -0.18, as "it is a special case, unique in the extremity of its exclusion of evolution from statewide science standards" (Lerner, 2000b, p.16). For example, it did not cover Darwin, biological evolution and any reference to the age of the earth. In this paper, I change this evolution score from -0.18 to 0 for ease of interpretability of regression results. All results using the original score of -0.18 for Kansas instead of 0 do not differ meaningfully (results available upon request). Iowa had no Science Standards in 2000 which is coded as missing. The District of Columbia is treated as a state throughout this paper. The evolution score was originally defined between 0 and 100, but I re-scale it by dividing it by 100, again for ease of interpretability. More information about the details of the scoring scheme are provided in Lerner (2000b, pp.10-17).

[11] This implies that reforms before the respective last reform are not taken into account in the analyses. In theory, ignoring these prior reforms merely cause attenuation bias as long as these prior reforms are uncorrelated with the timing of the last reform in a given state. To test for this explicitly, I perform a robustness check restricting the sample to students from states for which careful examination of academic articles, legal documents, and state education websites indicates that they had only one reform between 2000 and 2009, see Section 3.6.4 and Table A3.1 for more details.

[12] The standard high school curriculum typically features biology (the subject in which evolution is being taught) in the first year of high school.

[13] Figure A3.2 also depicts the evolution score levels in 2000 and 2009.

construction, the changes depend in part on the baseline level, in the sense that Science Standards which cover evolution very comprehensively in 2000 cannot expand the coverage by much until 2009, and vice versa. However, by identifying from changes within states I control for fixed differences between states. Overall, the evolution score changes are fairly well dispersed across the US, with each census region having at least one state in which the evolution coverage became more comprehensive, less comprehensive, and remained unchanged, respectively.

3.3.2 Micro Data

The following subsection describes the three micro-level datasets used in this paper. These repeated cross-sectional datasets are standardized and hence comparable across US states and cohorts, which makes them suitable for analyses with state and cohort fixed effects.

In all three datasets, I keep students in the sample who have no missings for basic controls variables and who enter high school after 1990 and before 2010 in my preferred sample cut. Hereby, I balance temporal proximity to the reform years with having sufficient years to estimate pre-trends and fixed effects credibly (and with statistical power in general). This approach also ensures not to identify from the adoption of the Next Generation Science Standards which started in 2013. The results of this paper do not hinge on this specific sample cut, as shown in robustness tests in Section 3.6.4.

NAEP: Evolution Knowledge in School

To estimate the effect of students' exposure to the teaching of evolution in high school on their knowledge about evolution by the end of high school, I link the evolution score with the restricted-use individual-level National Assessment of Educational Progress (NAEP). NAEP is a standardized student achievement test, measuring the knowledge of US students in various subjects since 1990. For this study, I use the NAEP test for science in grade 12 as it contains questions on evolution. Students are coded as exposed to the Science Standard in place in the year and state of their high school entry, assuming that they entered high school three years before taking the test in grade 12 in the same state.

The main outcome variable "evolution knowledge" is defined as the share of questions on evolution answered correctly. The nine categories of scientific knowledge on topics other than evolution are defined analogously. They serve as placebo outcomes in subsequent analyses and include topics such as "reproduction", "climate" or the "universe". In addition, the NAEP student surveys provide rich student-level control variables. They include, among others, variables measuring the socioeconomic status such as subsidized lunch status, parental education and home possessions.

The main sample only contains public school students, as Science Standards have never been binding for private schools. However, the latter serve as a placebo sample in robustness

checks. The main sample consists of more than 15,000 public school students who were asked at least one question on evolution. The descriptive statistics show that the average evolution score equals 0.65, implying that students in the sample were on average exposed to a "satisfactory" evolution coverage.[14] The mean of the main outcome variable "evolution knowledge" equals 0.32. The fact that not even one third of the questions on evolution are being answered correctly on average underscores the difficulty of the questions. Appendix A3.1 provides detailed tables of the descriptive statistics and raw correlations. It also presents sample questions, explains how the science questions are grouped into topical categories, and how missing observations are dealt with.

GSS: Evolution Approval in Adulthood

To estimate the effect of students' exposure to the teaching of evolution in high school on their approval of evolution in adulthood, I link the evolution score with the restricted-use individual-level General Social Survey (GSS). The GSS is a biennial cross-sectional survey which monitors societal change by interviewing a nationally representative sample of adults in the US since 1972. Since 2006, respondents are asked about their approval of evolution. The GSS also provides the state of residence at age 16 and the birth year. I assume that respondents entered high school in this state at age 14 and merge the evolution score for this state-year combination accordingly. Hence, I can link the approval of evolution of individuals in adulthood to the evolution coverage of the Science Standard they were exposed to as students, even if they migrated to other states after finishing school.

The main outcome variable "evolution approval" is based on the question "Human beings, as we know them today, developed from earlier species of animals. Is that true or false?".[15] The corresponding indicator variable is set to one if the answer "true" was given, and set to zero if any other answer option was reported such as "false", "don't know", or "no answer". The GSS also asks a broad range of questions on scientific topics other than evolution, and on religious, political and partisan attitudes. Other variables capturing different dimensions of the childhood environment are employed as control variables, including the religion a respondent was raised in.

The GSS is sampled from the entire US adult population regardless of type of school attendance. It does not allow to differentiate between public and private school attendance as the NAEP. Instead, one can estimate effects net of endogenous sorting across school types, including homeschooling. The estimation sample of individuals who were asked the question on evolution approval contains more than 1,800 individuals. The descriptive statistics show that 58 percent of the sample approve of evolution which is largely representative for evolu-

[14] Lerner (2000b) classifies evolution scores between 0.60 and 0.79 as "satisfactory".

[15] The words "human beings" are replaced by the word "elephants" for 10 percent of the questions on evolution approval in the sample. Table A3.7 shows that the results are robust to dropping these 10 percent from the sample.

tion approval in the US population at the time (Pew Research Center, 2009). More details on descriptive statistics, raw correlations and data background is provided in Appendix A3.1.

ACS: Occupational Choice

To estimate the effect of students' exposure to the teaching of evolution in high school on their probability to work in life sciences in adulthood, I link the evolution score with the individual-level IPUMS American Community Survey (ACS) (Ruggles et al., 2020). The ACS is a large-scale demographic survey which draws from a national random sample of the US population. Responding and providing correct information is required by US law. The ACS contains detailed information on the occupational field of the respondents. It also elicits the state and year of birth. I assume that students enter high school in this state at age 14 and merge the evolution score for this state-year combination accordingly.

Given that evolution is the fundamental theory of life sciences, the occupational field of primary interest in this study is life sciences. The main outcome variable "working in life sciences" is coded as an indicator variable equal to one if the respondent works in life sciences, and equal to zero otherwise. All other occupational fields are coded analogously. The ACS also allows to divide occupational fields into more fine-grained occupational subfields. The occupational field "life sciences" can be divided into the subfields "biology", "agriculture and food", "conservation and forestry" and "medical and other" for the purpose of subgroup analyses. Beyond sciences, I also analyze other occupational fields such as management, engineering and education. In total, there are 25 non-scientific occupational fields including one category for unemployed/not in the labor market which serve as placebo outcomes in robustness checks.

Like in the GSS, the ACS is sampled from the entire US population also including individuals who went to private school and homeschoolers. The estimation sample of individuals who are older than 18 years (i.e. who typically completed secondary education) consists of more than 6 million individuals. Further information, including descriptive statistics, is provided in Appendix A3.1.

3.4 Identification Strategy

All three analyses presented in this paper are based on the following two-way fixed effects (TWFE) model. The TWFE model exploits the different timing of reforms of the evolution coverage in Science Standards across states, and the fact that some of the reforms extended the coverage of evolution, while others reduced it, and a third group of states did not reform the evolution coverage. It compares outcomes of cohorts who went to high school in states where the evolution coverage was reformed with previous cohorts from the same states before the reforms, relative to how the outcomes of these cohorts changed in states who did not

reform at the time, after accounting for fixed differences between states and birth cohorts. The baseline parametric TWFE model is specified as follows:

$$Y_{istu} = \beta \cdot Evolution_Score_{st} + \gamma \cdot \mathbf{X_i} + \delta_s + \lambda_t + \theta_u + \epsilon_{istu} \tag{3.1}$$

where Y_{istu} is the outcome of interest of individual i, who entered high school in state s and year t, and completed the test or survey in year u. The treatment variable $Evolution_Score_{st}$ measures the intensity of the evolution coverage in the Science Standard in state s and year t. β is the parameter of interest capturing the effect on the outcome of being exposed to a very comprehensive coverage of evolution ($Evolution_Score_{st}$=1) as opposed to being exposed to no or a non-scientific/creationist coverage of evolution ($Evolution_Score_{st}$=0). The vector $\mathbf{X_i}$ contains the individual-level control variables. State fixed effects δ_s, birth cohort/high school entry cohort fixed effects λ_t, test/survey year fixed effects θ_u, and an error term complete the model. The standard errors are clustered at the state level which is conservative in this setting and accounts for the potential correlation of error terms across cohorts within states (Abadie et al., 2017; Athey and Imbens, 2021).

The baseline model addresses a range of concerns on the ability to estimate causal effects of the evolution coverage in Science Standards. One might be concerned that state-level differences in scientific, religious or political attitudes are correlated with the evolution coverage in Science Standards and affect scientific knowledge, attitudes as well as occupational choice. The state fixed effects absorb all differences in outcomes that are constant between states. In addition, one might be worried that national trends, such as attitudinal trends on scientific, religious or political topics, might erroneously appear as reform effects. To tackle this concern, the cohort fixed effects eliminate all national differences between cohorts.

A remaining concern are time-varying state-specific trends or shocks. For example, state-specific trends in human capital levels, or regional religiosity shocks induced by, say, church scandals may affect attitudes towards evolution differentially in different states. However, such factors only threaten the ability to estimate causal effects if they affect different high school entry cohorts differently. Many state-specific factors may be time-varying, but still affect adjacent cohorts similarly. This is the case, for example, if a church scandal occurring in a given year and state evokes similar reactions in adjacent cohorts. However, my empirical setup exploits cross-cohort variation within a narrow time window around the reforms, and identifies from reforms of the evolution coverage in State Standards that affect adjacent cohorts differently. Although reforms of Science Standards are generally applicable to all cohorts from the year of adoption onwards, the change in evolution coverage does typically only affect the high school entry cohort (and younger cohorts in the following years when they enter high school). This is true as the high school entry year is the year in which evolution

is typically being taught.[16] The state fixed effects capture such time-varying state-specific factors that affect students of different cohorts equally.

Moreover, I address concerns about time-varying state-specific trends or shocks that affect adjacent cohorts differently, but smoothly, by conducting robustness checks with state-specific time trends. For example, the trust in science among students could develop differently in the different states, but change smoothly across cohorts. The presented specification with linear and quadratic state-specific time trends is particularly demanding in terms of statistical power, as reform effects are only detectable as "jumps" from the cross-cohort trend. Showing that (at least the point estimates of) the main results hold in this specification reaffirms a causal interpretation of the presented findings.

In addition, the individual-level control variables take out observable differences between individuals that vary non-smoothly across states and cohorts. For example, controlling for the religion an individual was raised in ensures that outcomes across individuals are compared while holding constant their religion of upbringing.

Based on these insights, the TWFE model yields a causal effect if the parallel trend assumption holds. It requires that in the absence of Science Standard reforms, the change in outcomes in reforming states would have been the same as in non-reforming states. This is plausible given the institutional idiosyncrasies determining the exact reform timing, as discussed in Section 3.2. The following series of identification checks assesses its plausibility more formally.

I begin with conducting non-parametric event-study specifications, in which the reform of the evolution coverage in Science Standards in a given state and year is referred to as the “event”. In contrast to the baseline TWFE model, the event-study model can examine non-linear pre-reform trends in outcome variables. For example, evolution approval trending in the direction of estimated reform effects prior to the reform could indicate a bias from underlying trends in the data. Another advantage of the event-study models is that the timing of reform effects can be assessed by disentangling effects which occur directly at the time of the reforms from those which phase-in gradually after the reform. Specifically, I estimate the effect of the evolution coverage in Science Standards in year t_s on outcomes of students entering high school k years before and after the evolution coverage reform, as captured by the parameter vector β_k.

$$Y_{istu} = \sum_{k=-4}^{4} 1(t_{is} = t_s + k)\beta_k + \gamma \cdot X_i + \delta_s + \lambda_t + \theta_u + \epsilon_{istu} \tag{3.2}$$

[16] To the extent that evolution is also being taught in higher grade levels, the difference in exposure to the teaching of evolution between pre- and post-reform cohorts is overstated in my coding. Hence, I interpret the results as lower-bound estimates as parts of the cohorts coded as exposed to the pre-reform State Standard may be partially treated by post-reform State Standards.

Effects are estimated relative to the year of reform k=0.[17] The event-study estimations yield changes in outcomes induced by the *average* reform. This requires to run the event-study models separately for the subsets of states that expand and reduce the evolution coverage, respectively, because joint event-study models would cancel out effects from opposing reforms. Within each subset of states, the regression coefficients can be interpreted as changes in outcomes induced by the average reform in that subset (i.e. averaged over different evolution coverage changes within a subset).

As outlined above, a remaining concern to validity of the parametric and non-parametric analyses are time-varying state-specific shocks that affect adjacent cohorts differently, non-smoothly, and are not absorbed by the individual-level control variables. A series of placebo tests addresses this concern. In the first analysis, I test whether the coverage of evolution affects evolution knowledge (main outcome), but not knowledge on scientific topics other than evolution (placebo outcomes). In the second analysis, I test whether the coverage of evolution affects evolution approval in adulthood (main outcome), but not other scientific, religious or political attitudes (placebo outcomes). In the third analysis, I test whether the coverage of evolution affects the probability to work in life sciences (and in particular biology), but not the probabilities to work in non-scientific occupational fields. Null effects on placebo outcomes suggest that no previously uncontrolled scientific, religious or political shocks coincide with the timing of the reforms. They also demonstrate that the effects reported in this paper are narrowly tied to the topic of evolution, providing empirical support to the claim that the exact timing of reforms is driven by institutional idiosyncrasies instead of political changes. At the same time, they show that these outcomes themselves are not affected by the reforms.

Another placebo test makes use of a placebo sample of private school students for whom education standards have never been binding. One can test whether the reform effect on evolution knowledge is specific to public school students (main sample), but not detectable for private school students (placebo sample).

3.5 Results

This section shows in three steps that the evolution coverage in Science Standards affects the knowledge about evolution of students, the attitude on evolution in adulthood, and the probability to work in life sciences.

[17] To smooth the numbers of observations across years, the observations are grouped together to bins of 2 years. The bins at the beginning (end) of the domain additionally include the years prior to (following) the domain's starting (ending) year.

3.5.1 Evolution Knowledge in School

The first analysis demonstrates that the evaluated reforms affect what students learn about evolution in school. To that end, I regress the share of questions on evolution answered correctly in the 12^{th} grade NAEP science test on the evolution coverage in Science Standards and different sets of control variables. Column (1) of Table 3.1 displays the raw correlation without any control variables. The positive raw correlation could imply that being exposed to a comprehensive coverage of evolution increases students' knowledge about evolution (reform effect). However, it could also reflect that comparatively high average levels of evolution knowledge raise the probability that states adopt a Science Standard that covers evolution more comprehensively, for example because students might be less willing to accept creationist teaching (reverse causality). The positive raw correlation could also be driven by third variables such as parental education affecting both the probability that states adopt comprehensive Science Standards and the probability that students have knowledge about evolution (omitted variable bias).

To isolate the effect of the coverage of evolution in the Science Standard on evolution knowledge, I add different sets of control variables in columns (2)-(4). When adding student-level control variables in colum (2), or state and cohort fixed effects in column (3), or both the student controls and the fixed effects in column (4), the positive correlation persists and becomes even larger compared to the raw correlation. The full model in column (4) is the preferred specification as it exploits the reforms of Science Standards as a source of arguably exogenous variation by controlling for time-invariant differences between states, national differences between cohorts, time-varying state-specific shocks that affect adjacent cohorts similarly, as well as student level characteristics. It corresponds to the TWFE approach as specified in equation (3.1).

Regarding the main variable of interest, I find that being exposed to an evolution score of one, i.e. to a very comprehensive coverage of evolution, as opposed to an evolution score of zero, i.e. to no or a non-scientific/creationist coverage of evolution, increases the share of questions on evolution answered correctly by 5.8 percentage points. This effect is statistically significant at the 1 percent level. Given that students answer on average 32 percent of the questions on evolution correctly, the reported effect equals 18 percent of the sample mean.

Next, I hypothesize that the reform effect on evolution knowledge is disproportionately large for underprivileged students as they might rely more on schools to compensate for the lack of science exposure they receive from their parents and private environments. To begin, I note that variables typically associated with lack of privilege such as being Black (relative to being White) tend to predict knowledge about evolution negatively, see Table 3.1. Conversely, variables typically reflecting privilege such as having a computer at home tend to predict knowledge about evolution positively.

I perform subgroup analyses by student characteristics to assess whether reform effects are largest for underprivileged groups. Figure 3.2 shows that the effect of the evolution coverage on evolution knowledge is larger for Blacks than for any other racial/ethnic group. With regards to socioeconomic characteristics, results show that students without a computer at home and students receiving subsidized lunch benefit disproportionately from a comprehensive evolution teaching. In fact, the point estimate of 12.7 percentage points for students without a computer at home is the largest effect in this subgroup analysis. This finding seems plausible as having no computer at home does not only indicate a low socio-economic status, but also directly impedes access to online information sources through the internet. Apparently, students rely mostly on what is taught at school if access to independent information sources is restricted. I conclude from the findings of this subgroup analysis that a scientific content of education, materializing in this study in the form of a comprehensive evolution teaching, can act as a substitute for privilege in the production of student knowledge, materializing in this study in the form of evolution knowledge.

3.5.2 Evolution Approval in Adulthood

The second analysis shows that the teaching of evolution has lasting consequences on attitudes in adulthood, shedding light on the persistence of effects of scientific educational content. At the same time, it examines whether the effect on evolution knowledge translates into neutral settings in adulthood in which the scientifically correct answer is not encouraged. It could well be that students who are exposed to evolution content are able and willing to answer science exam questions correctly to gain points in an exam as the NAEP, but that they are not convinced of the correctness of evolution theory.

Table 3.2 presents the GSS results from regressions of evolution approval in adulthood on the evolution score in high school, conditional on different sets of control variables. The raw correlation in column (1) is positive and significant. When subsequently adding student-level controls and fixed effects, the effect becomes even larger. The estimate in the full model presented in column (4) shows that individuals who were exposed to an evolution score of one, as opposed to an evolution score of zero, are 33.3 percentage points more likely to approve of evolution in adulthood. This effect is highly significant and amounts to 57 percent of the sample mean, making it larger than the corresponding effect on evolution knowledge reported in section 3.5.1.

To benchmark the effect size relative to other determinants of attitudes, I calculate persuasion rates (DellaVigna and Gentzkow, 2010). The persuasion rate induced by a reform changing the evolution score from zero to one is defined as the average treatment effect on evolution approval divided by the share of students who do not approve of evolution in the entire sample.[18] The corresponding persuasion rate equals 79 percent. This is larger than the

[18] Another definition of the persuasion rate would require to divide the treatment effect of the average reform by the share of individuals who do not approve of evolution and who studied before the education standards

persuasion rates Cantoni et al. (2017) report for a Chinese school textbook reform on a range of outcomes.[19] It is also on the upper end of the persuasion rate distribution of media which includes rates from 3-8 percent (DellaVigna and Kaplan, 2007) to 65 percent (Enikolopov et al., 2011) for different media, settings and outcomes.

Regarding subgroups, the religion an individual was raised in gives rise to a particularly interesting heterogeneity analysis, given the large differences in attitudes on evolution and creationism between religious groups. I first document that having been raised as Evangelical is a large negative predictor of evolution approval in adulthood as compared to the other religious groups, see Table 3.2. Specifically, individuals raised as Evangelicals are 29 percentage points less likely to approve of evolution in adulthood compared to individuals being raised non-religiously, conditional on the other regressors. The predictive power for individuals raised as Mainline Protestants is substantially weaker. For Catholics, it is indistinguishable from those raised as non-religious.

The subgroup analysis depicted in Figure 3.3 shows that the reform effects are larger for individuals raised as Mainline Protestants relative to those from any other religious upbringing. Conversely, students are less susceptible to the effects of evolution teaching if they were raised in a religion with strong anti-evolution views on average like Evangelicals, or with strong pro-evolution views on average like those raised as non-religious. Mainline Protestants, with moderate evolution views on average, seem to be most open to change their attitude on evolution depending on the school curricula. Furthermore, the results show particularly strong reform effects for individuals that grew up in urban areas instead of rural areas. Moreover, the reform effects are largest for Blacks relative to all other racial/ethnic groups, which confirms the conclusions from the previous subsection.

3.5.3 Occupational Choice

The third analysis reveals that the teaching of evolution translates into real-world high-stakes outcomes beyond attitudinal outcomes. Specifically, I focus on occupational choice as one high-stakes life decision in which an individuals' attitudes, values and beliefs may be revealed. I hypothesize that exposure to evolution theory (and hence to the fundamental scientific theory about the existence of life) affects the probability that individuals choose to work in life sciences.

---

were reformed. However, compositional differences by states and cohorts between individuals who studied before and after the reforms would bias results. Similarly, calculating the persuasion rate based on predicting treated and untreated students' beliefs and subtracting the treatment effect from the treated students' beliefs as in Cantoni et al. (2017) is not feasible as most students are treated to some extent even before the reforms which then go in different directions with different intensities.

[19] They find the largest persuasion rates for the outcomes "Not investing in a bond" (50 percent persuasion rate) and "Trusting the local government" (47 percent persuasion rate).

Using the ACS, this analysis shows that being exposed to a more comprehensive teaching of evolution in school increases the probability to work in life sciences in adulthood, as presented in Table 3.3. The point estimate is significant and stable across specifications. The full model presented in column (4) shows that individuals who were exposed to an evolution score of one, as opposed to an evolution score of zero, are 0.035 percentage points more likely to work in life sciences as adults. This effect is significant at the 5 percent level but small in absolute terms due to the fact that few people work in life sciences in relation to the total US labor force. However, if expressed relative to the sample mean, the effect amounts to 23 percent.

The corresponding subgroup results by individual-level characteristics are in line with those from the previous subsections. Table A3.2 shows that the effect on the probability to work in life sciences is larger for females than for males, and for Blacks than for other racial/ethnic groups, if expressed relative to the respective subsample mean (although only the former is significantly different). The ACS does not provide more individual-level covariates, but one can conduct insightful subgroup analyses by the outcome variable, namely by the four subfields of life sciences. Figure 3.4 depicts a positive and highly significant effect of the reform on the probability to work in biology. It is large in relative size, amounting to more than 39 percent of the sample mean. For all other subfields of life sciences, the estimates show reform effects that are much smaller in size and not statistically different from zero. This subgroup pattern underpins that it is indeed the evolution teaching which drives reform effects, given the fundamental relevance of evolution for biology,[20] and given that evolution is being taught in biology.

3.6 Robustness

The presented TWFE estimations can be interpreted causally if the assumption of parallel trends holds, as described in section 3.4. The following subsections show non-parametric event-study graphs, placebo tests, robustness on time-varying treatment effects, and a battery of further specification checks including specifications with state-specific trends to assess the plausibility of the identifying assumption.

3.6.1 Event-Study Graphs

Non-parametric event-study graphs can assess pre-trends and the timing of reform effects. As shown below, there are no strong pre-trends for all three sets of analysis, particularly not in the direction of reform effects. Regarding reform effects, the changes in outcomes occur soon after reform adoption. The event-study models have to be conducted separately for the subsets of states that expand and reduce the evolution coverage, respectively, as joint

[20] This is illustrated by the well-known assertion by Dobzhansky (2013) that “nothing in biology makes sense except in the light of evolution”.

event-study models would cancel out effects from opposing reforms. In all event-study graphs, two years are grouped together to one bin to smooth the number of observations across bins.

Figure 3.5 displays no statistically significant pre-trend in evolution knowledge in both subsets. For the subset of states where the reform reduces the evolution coverage in Science Standards, there is, if at all, an upward trend in evolution knowledge before the reform. Against this positive insignificant pre-trend, evolution knowledge decreases significantly after the reform. For example, students entering high school one or two years after their state reduced the evolution coverage in the Science Standard display a reduction of 8 percentage point in the share of questions on evolution answered correctly in grade 12 relative to their counterparts who entered high school in the year of the reform or the year before. For the subset of states where the reform expands the evolution coverage in Science Standards, there is no pre-trend and an insignificant increase in evolution knowledge after the reform. Hence, evolution knowledge changes in the respective directions of the reforms in both cases, although the effect size and significance are larger for the subset of states reducing their evolution coverage.

Figure 3.6 depicts no significant pre-trends in evolution approval in adulthood in both subsets. The reform effects are, again, more pronounced for the subset of states in which the evaluated reforms reduce the evolution coverage. The same is true for the probability to work in life sciences, see Figure 3.7. In sum, the causal interpretation of the results presented in this paper is primarily based on the subset of states reducing the evolution coverage. Still, the effects on the subset of states expanding the evolution coverage go in the expected direction too.

3.6.2 Placebo Tests

A remaining threat to internal validity are state-specific shocks, events, or trends that affect adjacent cohorts differently and coincided with reforms of the evolution coverage in Science Standards and affect the respective outcomes. The following placebo tests are designed to assess this treat. I show below that neither changes in knowledge on non-evolution scientific topics at the end of high school, nor changes in non-evolution scientific, religious and political attitudes in adulthood, nor changes in the probabilities to work in non-scientific occupations appear as reform effects. These findings support (i) that reform coefficients do not reflect underlying shocks or trends and (ii) that the reforms themselves do not affect these outcomes. The fact that the reform effects are neatly tied to the topic of evolution in all three independent datasets and outcomes therein supports a causal interpretation of the results of this paper. In sum, the placebo tests provide empirical support for the theoretical assessment that the exact reform timing is determined by institutional idiosyncrasies and not by scientific, religious or political confounders.

Evolution Knowledge in School: As is visible in Table 3.4, there is no effect of the evolution coverage in Science Standards on student knowledge in any of the non-evolution scientific topics such as reproduction or climate. Column (11) also presents a regression in which the outcome variable is the average of the nine shares of questions answered correctly on

non-evolution scientific topics. The corresponding point estimate is insignificant and close to zero.

To rule out that shocks or events specific to evolution but not related to the Science Standards drive the main effect, I perform the main analysis on a placebo sample of students from private schools for whom the reforms were never compulsory. As shown in Table 3.5, the point estimate measuring the effect of the evolution coverage on evolution knowledge of private school students is very close to zero (although imprecisely estimated and therefore not significantly different from the point estimate of public school students). I conclude from this result that it is unlikely that there are shocks or events related to evolution coincident with the reform of evolution coverage in Science Standards, at least as long as they affect both public and private school students. This result also suggests that there are no spillovers from public school curricula to private school curricula. In addition, the main effect holds on a joint sample of both public and private school students addressing the concern that spurious selection of students or school curricula into (or out of) private schools coincidental to the reform drives the results.

Evolution Approval in Adulthood: Table 3.6 demonstrates that the evolution coverage does not affect non-evolution scientific outcomes in adulthood on topics such as radioactivity or viruses. This is true for each of the nine non-evolution scientific outcomes, and for the average of all nine outcomes. This finding can be interpreted as the adulthood equivalent of the placebo tests on non-evolution scientific outcomes measured at the end of high school shown above.

Table 3.7 shows that the evolution coverage has no effect on religious outcomes in adulthood. Religious outcomes include variables capturing (i) religious beliefs such as belief in God, (ii) religious belonging such as religious affiliation or churchgoing, and (iii) general religiosity.[21] There is no effect that is statistically different from zero on any religious outcome.

Table 3.8 demonstrates null effects of the reform on political outcomes. These outcomes comprise general political attitudes such as thinking of oneself as a Republican (as opposed to Democrat, Independent, or something else), political attitudes on specific topics typically regarded as controversial or partisan such as same-sex marriage, and preferences for governmental spending increases in areas such as alternative energy sources. There is no effect that is statistically different from zero on any political outcome.

Had there been, say, a negative coefficient on religiosity or political conservatism, it would be hard to disentangle whether this result was driven by confounding shocks or by the reforms. However, a null finding implies that neither confounding shocks nor reform effects are at work, because they plausibly operate in the same direction and do not offset each other. For

[21] The distinction between believing and belonging follows Barro and McCleary (2003) and McCleary and Barro (2019) who find in cross-country analyses that believing stimulates economic growth, while belonging tends to reduce economic growth at given levels of religious beliefs.

example, it would be implausible to assume that negative confounders, say, state-specific church scandals coincident with the reforms, reduce the coverage of evolution in Science Standards causing a negative effect on religiosity, while at the same time offsetting this negative effect by increasing religiosity through other channels.

Occupational Choice: Before turning to the placebo analysis on the probabilities to work in non-scientific occupations, I begin with contrasting the reform effect on life sciences with effects on other scientific occupational subfields. Table 3.9 only shows significant reform effects on natural sciences (life sciences and physical sciences, with the effect on life sciences being marginally larger if measured relative to its sample mean), but not on non-natural sciences (science technicians and social sciences). When looking at the overall reform effect on the probability to work in any scientific field, I find a positive and significant effect.

Placebo tests in which the other 25 non-scientific occupational fields serve as outcome variables allow for testing of whether reform effects are specific to working in scientific occupational fields. As presented in Table 3.10, the coverage of evolution has no significant effect on working in any non-scientific occupational field. There is also no effect on being unemployed or not being part of the labor market, implying that the reform does not cause selection into employment and the labor market.

3.6.3 Time-Varying Treatment Effects

Even in the absence of confounding trends or shocks, consistent estimation of reform effects requires homogeneity in treatment effects (Chaisemartin and D'Haultfœuille, 2020; Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Roth and Sant'Anna, 2021; Sun and Abraham, 2021). The treatment effect from the baseline TWFE model is a weighted average of all possible 2x2 difference-in-differences comparisons between treated and untreated groups as well as groups treated at different points in time (Goodman-Bacon, 2021). In settings with staggered treatment timing like in this study, time-varying treatment effects can bias results away from the true effect if already-treated students act as controls for later-treated students (negative weighting).

This concern can be addressed in a robustness check in which those 2x2 difference-in-differences comparisons in which already-treated students act as controls are excluded from the sample. Specifically, I implement the estimator by Callaway and Sant'Anna (2021) (CS estimator) which is designed to run with repeated cross-sectional data. Like with the event-study graphs, the CS estimator has to run separately for subsets of states that expand and reduce the evolution coverage, respectively, because joint CS estimations would cancel out effects from opposing reforms. Within each subset of states, the reform coefficients can be interpreted as changes in outcomes induced by the average reform in that subset (i.e. averaged over different evolution coverage changes within a subset).

Table A3.3 presents the CS estimator of the effect of the evolution coverage on evolution knowledge by the end of high school. For the subset of states reducing the evolution coverage, the estimate shows a reform effect of 5.6 percentage points. It is highly significant and very similar to the overall TWFE effect of 5.8 percentage points. For the subset of states expanding the evolution coverage, the CS estimator is only about half the size, and not significant. The CS estimator on evolution approval in adulthood is reported in Table A3.4. The CS estimator for the first subset of states amounting to 27.4 percentage points is relatively close to the 33.3 percentage points from the main TWFE model, although it does not met conventional levels of statistical significance. For the second subset of states, the effect is somewhat smaller in size and significance. Regarding the probability to work in life sciences, the CS estimator for the first subset of states is even marginally larger than the overall TWFE effect and comparable in terms of significance, see Table A3.5. The CS estimator for the second subset of states is still positive but insignificant and close to zero.

These findings underscore the conclusion from the event-study models that the causal interpretation of the findings presented in this paper rests on the sample of states reducing the evolution coverage. However, the results from the group of states expanding the evolution coverage go in the expected direction too.

3.6.4 Further Robustness Checks

This subsection covers a battery of further robustness checks. The first test replicates the main analysis on a subset of reforms which themselves can arguably be regarded as-good-as-random (and not only their specific timing). This subset contains reforms in states in which the governor decides about the members of the State Board of Education, and in which the governor ruling at the time of the reform adoption won the previous election by a small margin. In these states, the outcome of the election and hence the political direction of the Boards of Education and their reforms is somewhat arbitrary. Although the set of states with close pre-reform gubernatorial elections reduces the sample size by about two thirds, the reform effects on evolution knowledge are robust (see column (1) of Table A3.6). The same is true for analogous analyses on evolution approval in adulthood (see column (1) of Table A3.7) and on the probability to work in life sciences (with the latter being estimated less precisely, see column (1) of Table A3.8). These findings lend empirical support to a causal interpretation of the presented estimation results, even if it was not true that institutional idiosyncrasies were quasi-randomizing the reform timing.

Another, more direct, way to control for political changes is the inclusion of state-by-year controls for the political affiliation of the governor ruling in the state and year of the respective individuals' high school entry. As reported in column (2) of Tables A3.6, A3.7, and A3.8, respectively, this test yields robust results throughout the three analyses both in terms of size and significance.

Adding state-specific time trends as control variables to the baseline TWFE model constitutes another way to assess robustness. These trends explicitly account for time-varying state-specific shocks that affect adjacent cohorts differentially, but smoothly. As is visible in column (3) of the three Tables listed above, the levels of significance tend to decrease in this demanding specification, while the point estimates largely hold and partly become even larger.

Another robustness check reduces the sample to states which had only one reform event between 2000 and 2009 based on careful examination of academic articles, legal documents, and state education websites. As shown in column (4) of the three Tables listed above, the results are largely robust and partly even more pronounced.

In addition, the results hold if the observation period of the main sample is defined differently. As reported in columns (5) and (6) of the Tables listed above, the results are robustness to sample definitions with fewer pre-reform cohorts, with the earliest cohorts entering high school in 1995 and 2000, respectively. Moreover, the results do not hinge on the precise coding of the outcome variables (for this test, the column numbers depends on the analysis, see footnotes of the three Tables listed above for more information). For example, the results are robust to coding those individuals who do not know how to answer the question on evolution approval as a missing observation instead of non-approving. There are also corresponding results for the analysis on evolution knowledge, but not for the probability to work in life sciences as the latter has no such outcome category. The remaining columns of the three Tables listed above show that all results are robust to conducting logit and probit specifications, and to dropping missing observations of control variables instead of imputing them.

Lastly, the interpretation of the results does not change when transforming the treatment variable to indicator variables. Specifically, the first (second) indicator variable is set to one if the evolution score is larger than 0.1 (0.2), and zero otherwise. The seven other indicator variables are coded accordingly. This coding eliminates a substantial amount of treatment variation, but allows to assess which domain of the evolution score distribution is particularly important for the production of evolution knowledge, evolution approval, and the probability to work in life sciences. Tables A3.9, A3.10, and A3.11 show that most domains of the evolution score distribution are important for the production of outcomes with the exception of the highest value.

3.7 Conclusion

This paper shows that the content of science education has lasting effects on students by affecting their knowledge, attitudes and choices. To demonstrate this, the paper focuses on the teaching of evolution theory in the US. Exploiting institutional idiosyncrasies in the timing of reforms of the evolution coverage in US State Science Education Standards, I first document that the teaching of evolution causally affect student's knowledge about evolution at the end of high school. Second, I show that the teaching of evolution shapes the attitudes

on evolution of exposed students in adulthood, while non-evolution scientific, religious and political attitudes remain unaffected. The null finding on religious outcomes speaks against concerns in the policy debate at the time that expanding the teaching of evolution may reduce students' religious convictions. Third, I demonstrate that the teaching of evolution impacts high-stakes life decisions as exemplified by occupational choice. All three sets of results are of direct economic importance given the effects on individual and societal outcomes of scientific knowledge (Hanushek and Woessmann, 2008), scientific attitudes (Brzezinski et al., 2021), and of working in STEM occupations (Peri et al., 2015). Consensus on topics such as evolution could also reduce societal polarization and its associated costs (Alesina et al., 2020). In sum, I conclude from the three sets of findings presented in this paper that science education has lasting effects on students.

To illustrate the effect sizes, I calculate changes in outcomes that one would expect to observe if all states adopted Science Standards with a highly comprehensive evolution coverage relative to the average coverage in the sample. Linear extrapolation of the presented estimation results suggests that the evolution approval in the US population would increase by 20 percent of the sample mean in such a scenario. Analogously, the number of adults working in life sciences would increase by 8 percent of the sample mean, and in the subfield of biology by 13 percent of the sample mean.

This paper shows that the content of education standards is relevant for individuals in the short- and long-run. This conclusion challenges the notion that education standards have no meaningful impact on students which is prevalent in the academic and political debate. It has been argued that there is little room for education standards to affect teaching in reality due to the dominating role of other factors such as the teachers' own ideology for curriculum design in school (Moore et al., 2003a; Loveless, 2021). Still, legal pressures on school districts to follow education standards, the reflection of the content of education standards in textbooks, as well as the gradual expansion of standardized testing covering the content of education standards have arguably incentivized teachers to follow education standards. The analyses presented in this paper empirically demonstrate that they indeed affect what students learn.

More broadly, this paper shows that the content of school curricula and instruction lastingly shapes students. This is true even for a topic like evolution that is highly charged in political and societal debates. Despite its fundamental relevance for and overwhelming acceptance in science, people have strong partisan views on it. These views are likely determined by a multitude of factors. Still, what schools teach has long-run effects on individuals' fundamental views and translates into high-stakes choices.

Beyond the evolution content of US State Science Education Standards evaluated in this paper, the findings presented here might imply that other US education policies increasing the time teachers spend on teaching evolution could have created analogous effects.[22] Ex-

[22] Between 2007 and 2019, the average number of hours a high school biology teachers in U.S. public schools spends on teaching human evolution almost doubled from 4.1 to 7.7 class hours (Plutzer et al., 2020).

amples of such policies include the adoption of the Next Generation Science Standards, and improvements in pre-service teacher education (Plutzer et al., 2020).

Beyond the US, the question of whether evolution should be taught in school is controversial in many other countries too.[23] If such countries expanded their evolution teaching, one might observe analogous effects on the scientific knowledge, attitudes, and related life decisions of their citizens.

Beyond the topic of evolution, the findings of this paper might also be relevant more broadly for further topics of science teaching, such as vaccinations, climate change or the trust in science in general. It is up to future research to show that explicitly.

[23] This can be illustrated by the headlines "Turkey's new school year: Jihad in, evolution out" by the BBC (Altunas, 2017), "Israeli schools largely avoid teaching evolution" by the Times of Israel (Staff, 2018), and "Indian education minister dismisses theory of evolution" by the Guardian (Safi, 2018).

Figures and Tables

Table 3.1: Effect of evolution coverage in Science Standards on evolution knowledge in school

	Evolution Knowledge			
	(1)	(2)	(3)	(4)
Evolution Score	0.036*	0.039***	0.069*	0.058**
	(0.018)	(0.011)	(0.028)	(0.019)
Female		-0.003		-0.004
		(0.008)		(0.008)
Race/Ethnicity: Black		-0.084***		-0.082***
		(0.007)		(0.007)
Race/Ethnicity: Hispanic		-0.051***		-0.048***
		(0.009)		(0.009)
Subsidized Lunch		-0.012		-0.011
		(0.006)		(0.006)
Parental Education: Graduated High School		-0.009		-0.010
		(0.011)		(0.012)
Parental Education: Some education after High School		0.002		0.002
		(0.011)		(0.011)
Parental Education: Graduated College		0.023*		0.021
		(0.010)		(0.011)
Computer at Home		0.011		0.022**
		(0.007)		(0.007)
State FEs	NO	NO	YES	YES
Birth Year FEs	NO	NO	YES	YES
Other Controls	NO	YES	NO	YES
Mean of Dep. Var.	0.32	0.32	0.32	0.32
Std. Dev. of Dep. Var.	0.42	0.42	0.42	0.42
Adj. R-squared	0.001	0.043	0.015	0.049
Observations	15530	15520	15530	15520

Notes: Dependent variable: Share of questions about evolution answered correctly. Other Controls: Indicator variables for asian (race/ethnicity) other (race/ethnicity), English language learner status, disability status, parental education, home possessions (books), and test year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table 3.2: Effect of evolution coverage in Science Standards on evolution approval in adulthood

	Evolution Approval			
	(1)	(2)	(3)	(4)
Evolution Score	0.108**	0.089**	0.205	0.333**
	(0.040)	(0.033)	(0.115)	(0.107)
Female		-0.053*		-0.050*
		(0.022)		(0.022)
Race/Ethnicity: Black		-0.158***		-0.149***
		(0.038)		(0.040)
Race/Ethnicity: Hispanic		-0.100*		-0.091
		(0.044)		(0.056)
Raised in Rural Area		-0.014		-0.003
		(0.024)		(0.025)
Raised as Protestant: Mainline		-0.141***		-0.121**
		(0.035)		(0.035)
Raised as Protestant: Evangelical		-0.302***		-0.290***
		(0.046)		(0.047)
Raised as Catholic		0.018		0.019
		(0.037)		(0.040)
State FEs	NO	NO	YES	YES
Birth Year FE	NO	NO	YES	YES
Other Controls	NO	YES	NO	YES
Mean of Dep. Var.	0.58	0.58	0.58	0.58
Std. Dev. of Dep. Var.	0.49	0.49	0.49	0.49
Adj. R-squared	0.005	0.088	0.038	0.107
Observations	1,812	1,801	1,812	1,801

Notes: Dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know). Other Controls: Indicator variables for white (race/ethnicity; omitted category) other (race/ethnicity), parents born abroad, parental education, having lived with parents in adolescence, religion raised in (Indicator variables for mainline protestantism, evangelical protestantism, catholicism (all reported here), no religion (omitted category), judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table 3.3: Effect of evolution coverage in Science Standards on probability to work in life sciences

	Life Sciences			
	(1)	(2)	(3)	(4)
Evolution Score	0.039*	0.035*	0.035*	0.035*
	(0.018)	(0.013)	(0.014)	(0.014)
Female		0.014*		0.013*
		(0.006)		(0.006)
Race/Ethnicity: Black		-0.127***		-0.115***
		(0.007)		(0.006)
Race/Ethnicity: Hispanic		-0.106***		-0.085***
		(0.008)		(0.008)
State FEs	NO	NO	YES	YES
Birth Year FEs	NO	NO	YES	YES
Other Controls	NO	YES	NO	YES
Mean of Dep. Var.	0.15	0.15	0.15	0.15
Std. Dev. of Dep. Var.	3.84	3.84	3.84	3.84
Adj. R-squared	0.000	0.000	0.000	0.001
Observations	6,460,650	6,460,650	6,460,650	6,460,650

Notes: Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). Other controls: Indicator variables for asian (race/ethnicities), other (race/ethnicities), multiple (race/ethnicities), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table 3.4: Placebo tests: Effect of evolution coverage in Science Standards on knowledge about non-evolution scientific topics

	Main Outcome:	Placebo Outcomes (Knowledge on the following Non-Evolution Scientific Topic):									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Evolution	Motion	Matter and Mass	Energy	Reproduction	Climate	Pollution	Earth	Tectonics	Universe	Average
Evolution Score	0.058**	-0.006	0.013	-0.014	0.024	-0.026	-0.014	0.039	0.009	0.003	0.006
	(0.019)	(0.045)	(0.047)	(0.024)	(0.028)	(0.046)	(0.059)	(0.023)	(0.020)	(0.051)	(0.012)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.32	0.51	0.30	0.38	0.38	0.39	0.15	0.41	0.17	0.31	0.35
Std. Dev. of Dep. Var.	0.42	0.43	0.43	0.43	0.42	0.39	0.27	0.42	0.27	0.42	0.28
Adj. R-squared	0.049	0.090	0.140	0.110	0.100	0.073	0.077	0.140	0.312	0.049	0.168
Observations	15520	9510	17000	22910	18610	19080	4770	13710	6730	6260	32850

Notes: Dependent variables: Shares of questions answered correctly about scientific topics indicated in the column headers. Controls: Indicator variables for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate Indicator variables for computer and books), and test year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table 3.5: Placebo tests: Effect of evolution coverage in Science Standards on evolution knowledge in private schools

	Evolution Knowledge		
	(1) Only Public School Students	(2) Only Private School Students	(3) Overall
Evolution Score	0.058**	0.003	0.046*
	(0.019)	(0.062)	(0.018)
State FEs	YES	YES	YES
Birth Year FEs	YES	YES	YES
Controls	YES	YES	YES
Mean of Dep. Var.	0.32	0.43	0.34
Std. Dev. of Dep. Var.	0.42	0.38	0.41
Adj. R-squared	0.049	0.045	0.056
Observations	15520	3160	18680

Notes: Regressions by students' school type as indicated in the column headers. Dependent variable: Share of questions about evolution answered correctly. Controls: Indicator variables for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate Indicator variables for computer and books), and test year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table 3.6: Placebo tests: Effect of evolution coverage in Science Standards on non-evolution scientific topics

	Main Outcome:	Placebo Outcomes: Non-Evolution Scientific Topics									
	(1) Evolution	(2) Earth	(3) Radioactivity	(4) Reproduction	(5) Lasers	(6) Electrons	(7) Viruses	(8) Big Bang	(9) Tectonics	(10) Sun	(11) Average
Evolution Score	0.333**	0.000	-0.125	0.192	0.042	-0.133	0.175	-0.191	-0.179	-0.266	-0.053
	(0.107)	(0.091)	(0.138)	(0.107)	(0.181)	(0.144)	(0.158)	(0.113)	(0.091)	(0.164)	(0.057)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.58	0.88	0.66	0.62	0.47	0.57	0.51	0.45	0.85	0.79	0.65
Std. Dev. of Dep. Var.	0.49	0.32	0.47	0.48	0.50	0.49	0.50	0.50	0.36	0.41	0.22
Adj. R-squared	0.107	0.051	0.074	0.035	0.091	0.038	0.092	0.113	0.054	0.090	0.158
Observations	1,801	1,800	1,797	1,747	1,799	1,800	1,801	1,796	1,801	1,801	1,801

Notes: Dependent variables: Questions answered correctly about scientific topics indicated in the column headers. Controls: Indicator variables for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (Indicator variables for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table 3.7: Placebo tests: Effect of evolution coverage in Science Standards on religious outcomes

	Main Outcome:	Religious Placebo Outcomes: Believing					Religious Placebo Outcomes: Belonging and Activities					Religious Placebo Outcomes: Overall			Religious Placebo Outcomes: Average
	(1) Evolution	(2) God	(3) Bible	(4) Afterlife	(5) Rebirth	(6) Strong Believer	(7) Religious Affiliation	(8) Church-going	(9) Church Activities	(10) Personal Prayer	(11) Missionize	(12) Spiritual Person	(13) Religious Person	(14) Fundamentalist	(15) Average
Evolution Score	0.333**	-0.021	-0.009	-0.023	-0.097	-0.091	0.099	0.102	-0.072	-0.108	-0.028	0.060	-0.206	0.015	-0.029
	(0.107)	(0.096)	(0.114)	(0.128)	(0.123)	(0.150)	(0.109)	(0.120)	(0.091)	(0.113)	(0.073)	(0.101)	(0.162)	(0.129)	(0.068)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.58	0.87	0.72	0.73	0.34	0.32	0.70	0.35	0.17	0.65	0.41	0.56	0.43	0.24	0.50
Std. Dev. of Dep. Var.	0.49	0.33	0.45	0.45	0.47	0.47	0.46	0.48	0.38	0.48	0.49	0.50	0.50	0.43	0.28
Adj. R-squared	0.107	0.104	0.086	0.029	0.134	0.086	0.163	0.081	0.040	0.149	0.130	0.061	0.077	0.244	0.201
Observations	1,801	1,797	1,794	1,797	1,796	1,783	1,799	1,801	1,801	1,798	1,801	1,801	1,799	1,718	1,801

Notes: Dependent variables: Main outcome and religious outcomes indicated in the column headers. Controls: Indicator variables for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (Indicator variables for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table 3.8: Placebo tests: Effect of evolution coverage in Science Standards on political outcomes

	Main Outcome:	Political Placebo Outcomes: General		Political Placebo Outcomes: In favor of:							Political Placebo Outcomes: Increase governmental spending for:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Evolution	Repub-lican	Conser-vative	Prayer in Public Schools	Sex Education in Public Schools	Same-Sex Marriage	Abor-tion	Marijuana Legali-zation	Capital Punish-ment	Gun Control	Immi-gration	Environ-ment	Alternative Energy Sources	Educa-tion	Scientific Research	Reducing Income Differences	Assistance to the Poor	Conditions of Blacks
Evolution Score	0.333**	0.028	-0.018	-0.061	0.085	0.084	-0.257	-0.006	-0.101	-0.013	-0.082	-0.062	0.045	-0.047	-0.116	0.137	0.148	-0.044
	(0.107)	(0.114)	(0.132)	(0.104)	(0.065)	(0.225)	(0.188)	(0.153)	(0.162)	(0.135)	(0.141)	(0.105)	(0.146)	(0.104)	(0.138)	(0.100)	(0.101)	(0.116)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.58	0.28	0.26	0.45	0.94	0.64	0.45	0.55	0.61	0.70	0.16	0.71	0.59	0.81	0.37	0.54	0.49	0.36
Std. Dev. of Dep. Var.	0.49	0.45	0.44	0.50	0.25	0.48	0.50	0.50	0.49	0.46	0.36	0.45	0.49	0.39	0.48	0.50	0.50	0.48
Adj. R-squared	0.107	0.058	0.024	0.087	0.048	0.110	0.028	0.079	0.063	0.060	-0.001	0.018	0.011	0.027	0.012	0.001	0.060	0.095
Observations	1,801	1,792	1,791	1,200	1,200	1,056	1,059	1,336	1,788	1,063	1,174	1,801	1,195	1,801	1,799	1,337	1,798	1,788

Notes: Dependent variables: Main outcome and political outcomes indicated in the column headers. Controls: Indicator variables for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (Indicator variables for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table 3.9: Effect of evolution coverage in Science Standards on probability to work in broader scientific occupational subfields

	(1) Life Sciences	(2) Physical Sciences	(3) Social Sciences	(4) Science Technicians	(5) Overall: All Sciences
Evolution Score	0.035*	0.042*	0.031	-0.027	0.081*
	(0.014)	(0.018)	(0.028)	(0.053)	(0.037)
State FEs	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.15	0.22	0.16	0.32	0.85
Std. Dev. of Dep. Var.	3.84	4.68	4.03	5.62	9.16
Adj. R-squared	0.00064	0.00083	0.00096	0.00073	0.00166
Observations	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650

Notes: Dependent variable: Probability to work in occupational field indicated in the column header (multiplied by 100 for interpretability). Controls: Indicator variables for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table 3.10: Placebo test: Effect of evolution coverage in Science Standards on probability to work in non-scientific occupational fields

	Main Outcome:	Placebo Outcomes: Non-Scientific Occupational Fields												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Sciences	Manage-ment	Business Operations	Finance	IT	Engi-neering	Social	Legal	Educa-tion	Arts	Health-care	Health-care Support	Protec-tive Services	Food
Evolution Score	0.081*	0.281	0.127	0.120	0.115	0.023	0.133	0.145	-0.015	0.074	-0.053	-0.045	0.035	0.125
	(0.037)	(0.191)	(0.099)	(0.079)	(0.080)	(0.068)	(0.097)	(0.078)	(0.241)	(0.111)	(0.189)	(0.182)	(0.201)	(0.386)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.85	5.44	1.87	1.63	1.97	1.39	1.39	0.82	5.58	2.04	4.33	2.63	2.13	7.88
Std. Dev. of Dep. Var.	9.16	22.68	13.54	12.67	13.89	11.69	11.70	9.01	22.96	14.15	20.36	15.99	14.43	26.94
Adj. R-squared	0.00166	0.01926	0.00489	0.00421	0.01092	0.00774	0.00417	0.00382	0.02060	0.00198	0.02620	0.01644	0.00655	0.01962
Observations	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650

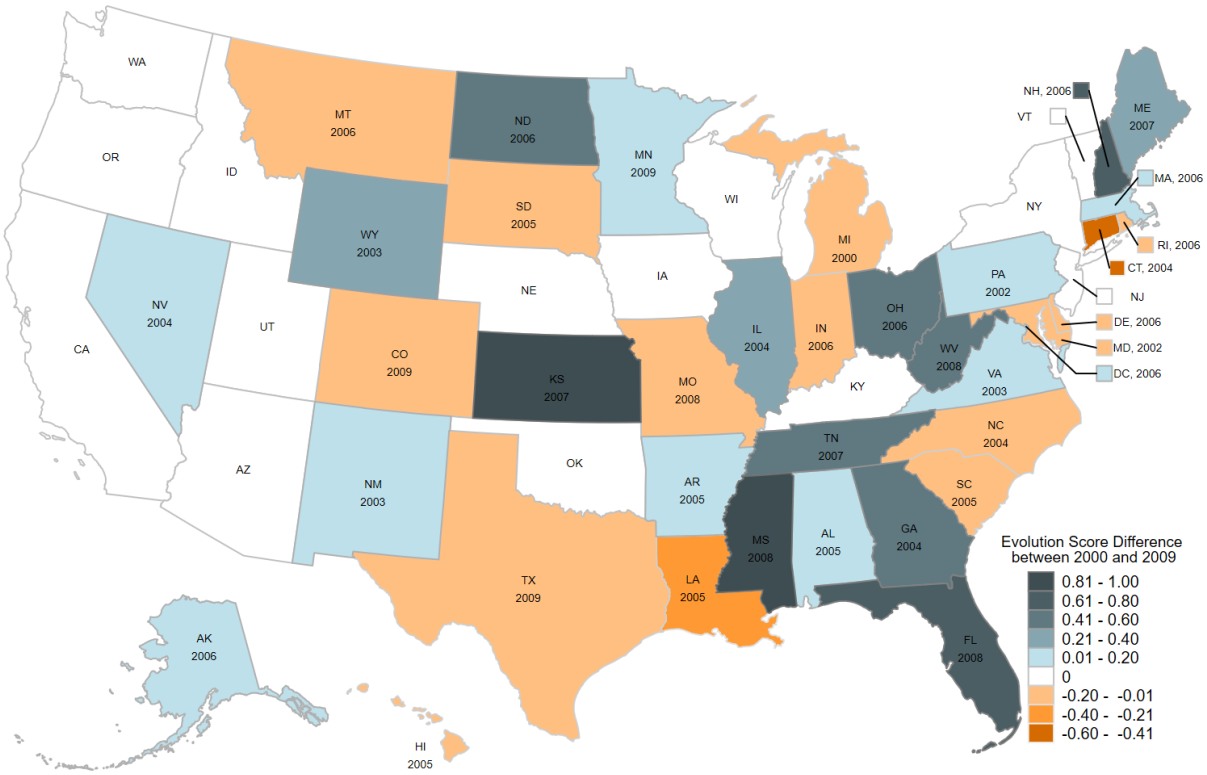
Notes: Dependent variable: Probability to work in occupational field indicated in the column header (multiplied by 100 for interpretability). Controls: Indicator variables for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table 10 (continued) – Placebo test: Effect of evolution coverage in Science Standards on probability to work in non-scientific occupational fields

	Placebo Outcomes: Non-Scientific Occupational Fields											
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
	Buildings	Personal Care	Sales	Office	Farming	Construc-tion	Extrac-tion	Installa-tion	Produc-tion	Transporta-tion	Armed Forces	Unemployed/ Not in Labor Market
Evolution Score	0.073	-0.078	-0.502	-0.059	-0.011	-0.227	0.058	0.060	-0.171	0.190	0.046	-0.525
	(0.283)	(0.284)	(0.267)	(0.488)	(0.111)	(0.249)	(0.047)	(0.117)	(0.220)	(0.329)	(0.080)	(1.778)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	2.84	3.96	11.42	13.25	0.66	4.45	0.21	2.80	4.87	5.35	0.75	9.48
Std. Dev. of Dep. Var.	16.61	19.51	31.81	33.90	8.07	20.63	4.60	16.50	21.53	22.50	8.62	29.30
Adj. R-squared	0.00622	0.01477	0.01027	0.02388	0.00547	0.04532	0.00463	0.02608	0.01894	0.02964	0.00522	0.03058
Observations	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650

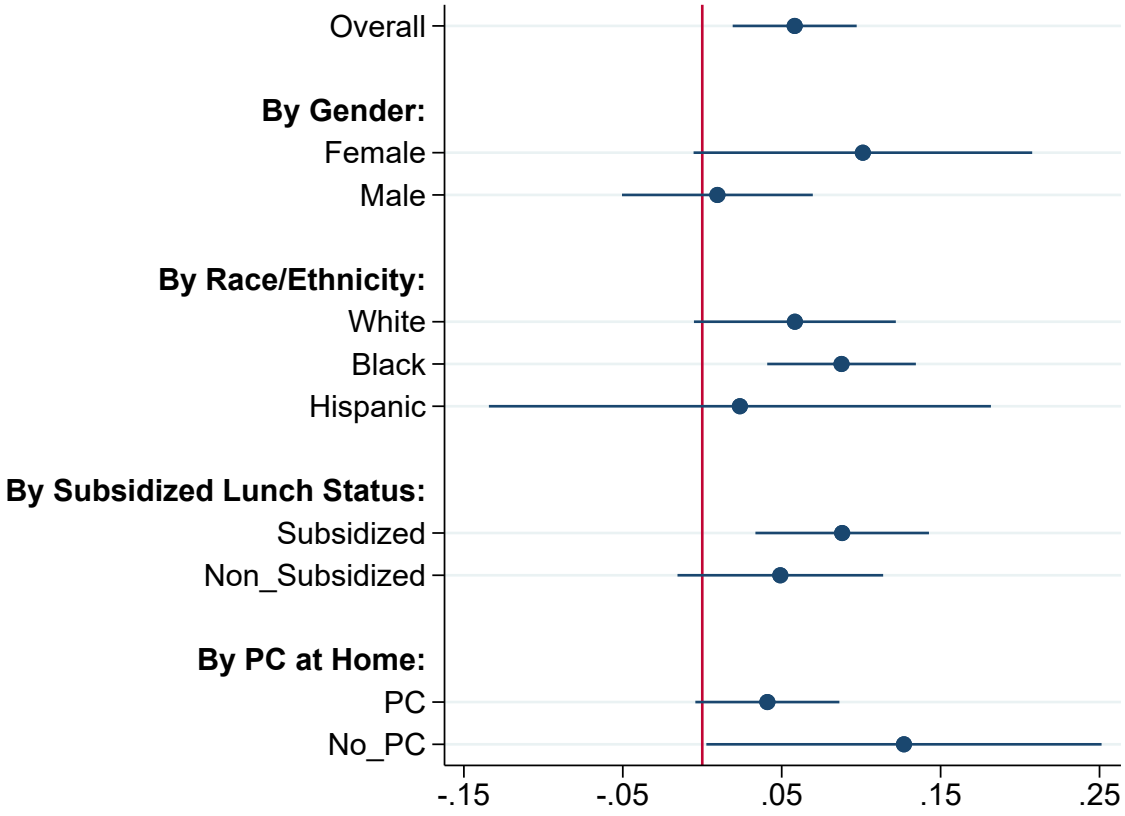
Notes: Dependent variable: Probability to work in occupational field indicated in the column header (multiplied by 100 for interpretability). Controls: Indicator variables for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Figure 3.1: US map of evolution score difference between 2000 and 2009



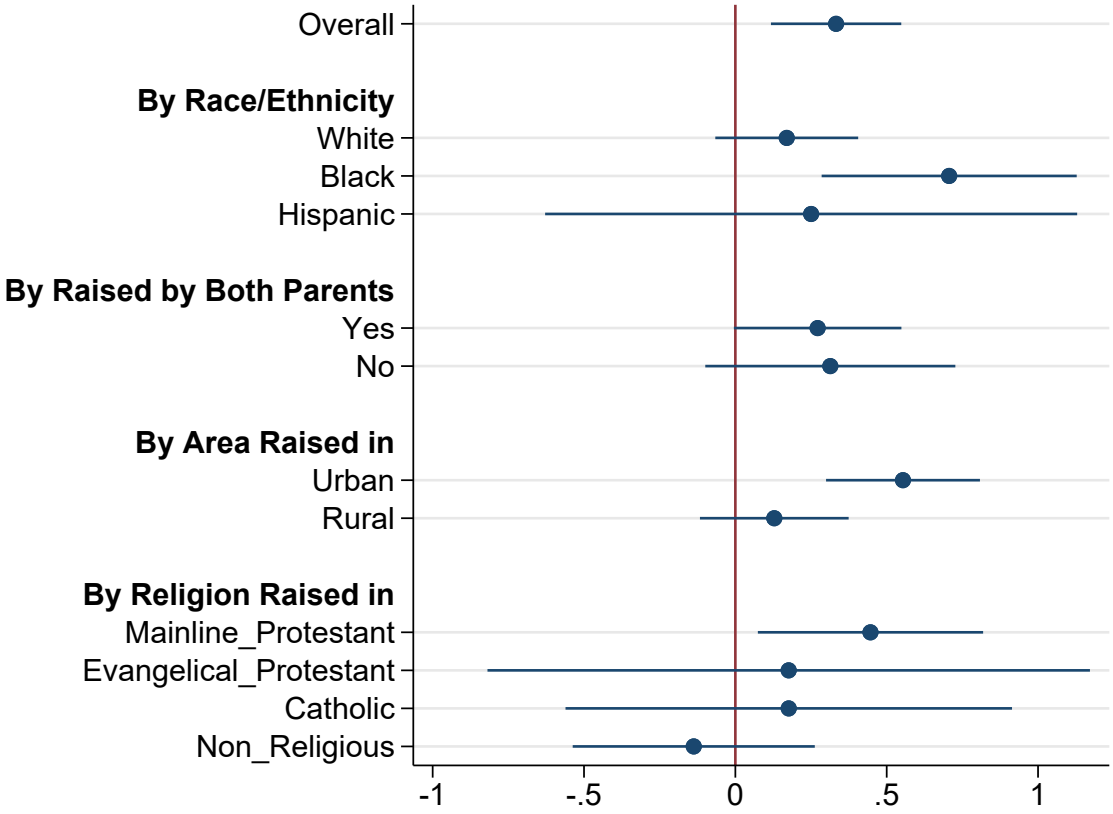
Note: Map depicts the evolution score difference, which I define as the evolution score of 2009 minus the evolution score of 2000. A positive (negative) difference implies an increase (decrease) in the evolution score between 2000 and 2009, as indicated by blue (orange) coloring. White coloring indicates no change of the evolution score between 2000 and 2009. The years reported below the two-letter state codes mark the respective reform years. A list of the evolution score differences and reform years underlying this map is provided in Table A3.1. Data source: Lerner (2000b), Mead and Mates (2009)

Figure 3.2: Effect of evolution coverage in Science Standards on evolution knowledge in school, by subgroups



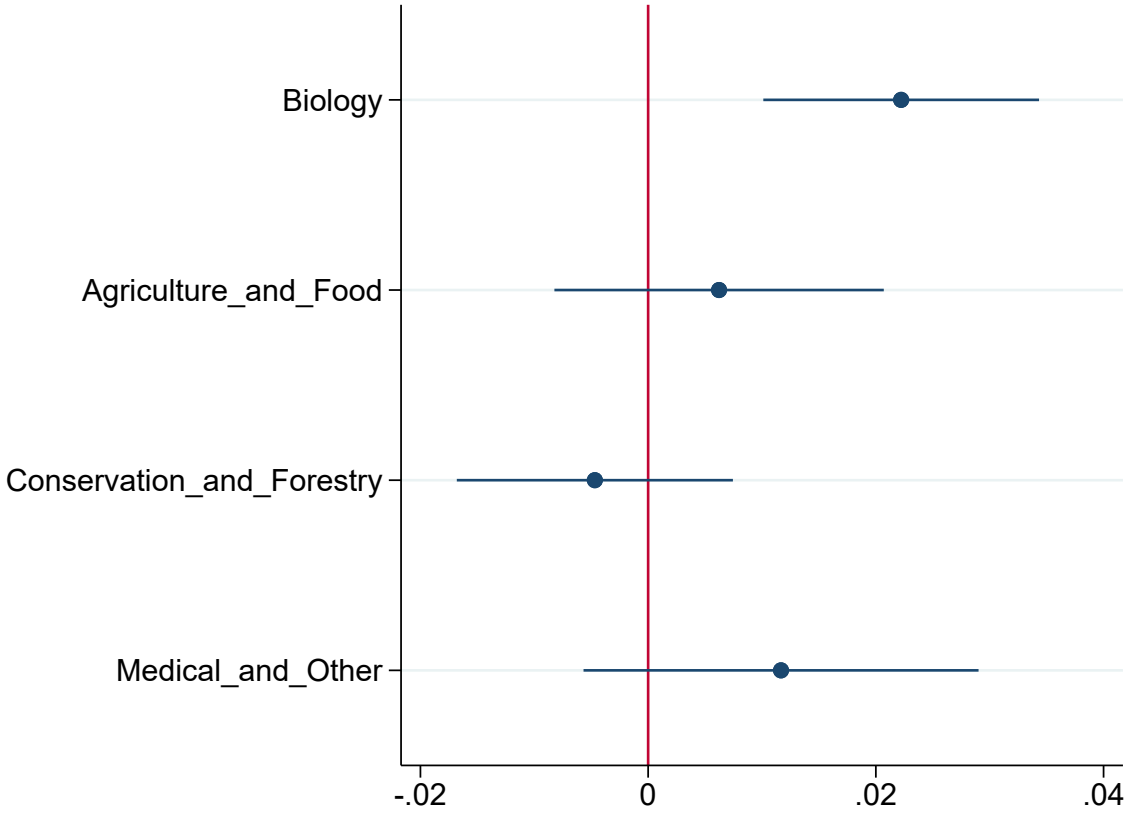
Note: Figure displays Effect of evolution coverage in Science Standards on share of questions about evolution answered correctly, by individual subgroup as indicated in rows. Controls: Indicator variables for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate Indicator variables for computer and books), and fixed effects for state, birth year, and test year. Standard errors clustered at the state level. 95% confidence intervals displayed. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Figure 3.3: Effect of evolution coverage in Science Standards on evolution approval in adulthood, by subgroups



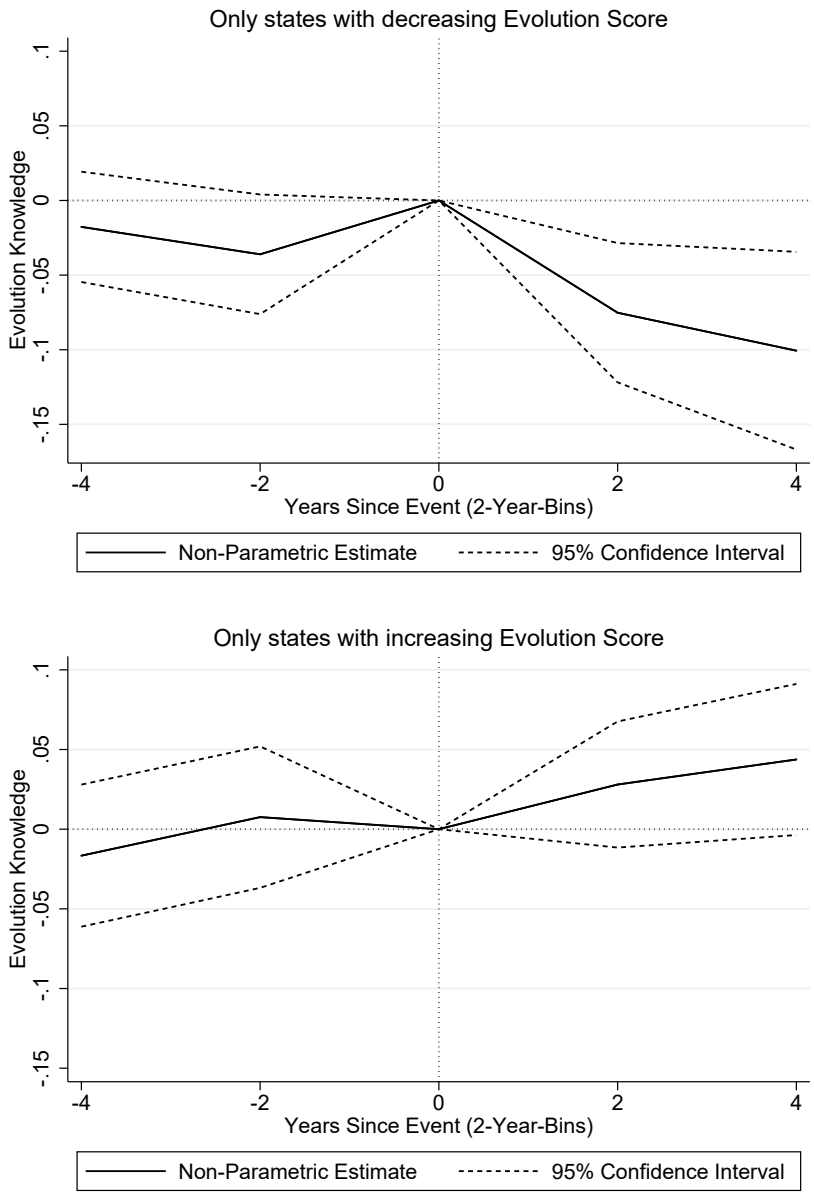
Note: Figure displays Effect of evolution coverage in Science Standards on approval of evolution in adulthood ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know), by individual subgroup as indicated in rows. Controls: Indicator variables for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (Indicator variables for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and fixed effects for state, birth year, and test year. Standard errors clustered at the state level. 95% confidence intervals displayed. Data source: General Social Survey

Figure 3.4: Effect of evolution coverage in Science Standards on probability to work in life sciences, by subfields of life sciences



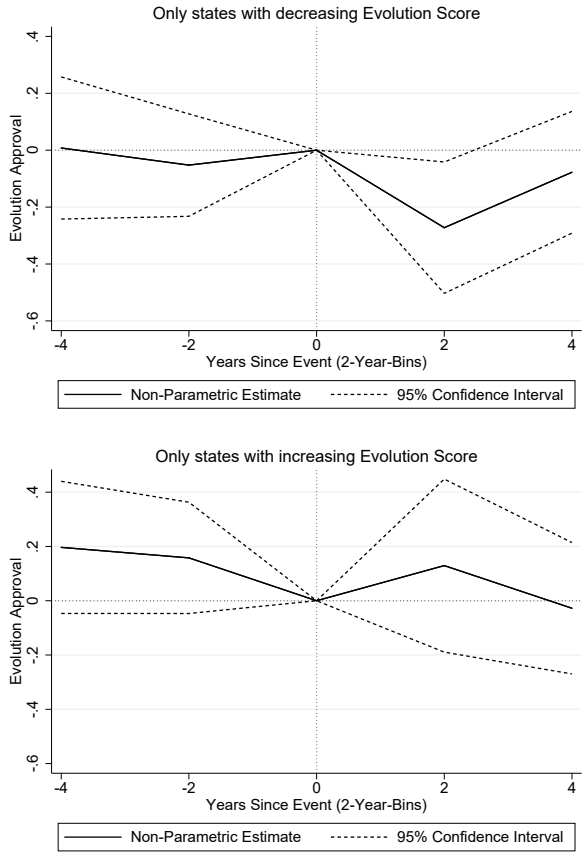
Note: Figure displays Effect of evolution coverage in Science Standards on probability to work in life sciences, by subfields of life sciences as indicated in rows (multiplied by 100 for interpretability). Controls: Indicator variables for gender, races/ethnicities and fixed effects for state, birth year, and test year. Standard errors clustered at the state level. 95% confidence intervals displayed. Data source: American Community Survey

Figure 3.5: Event-study graphs: Evolution knowledge in school



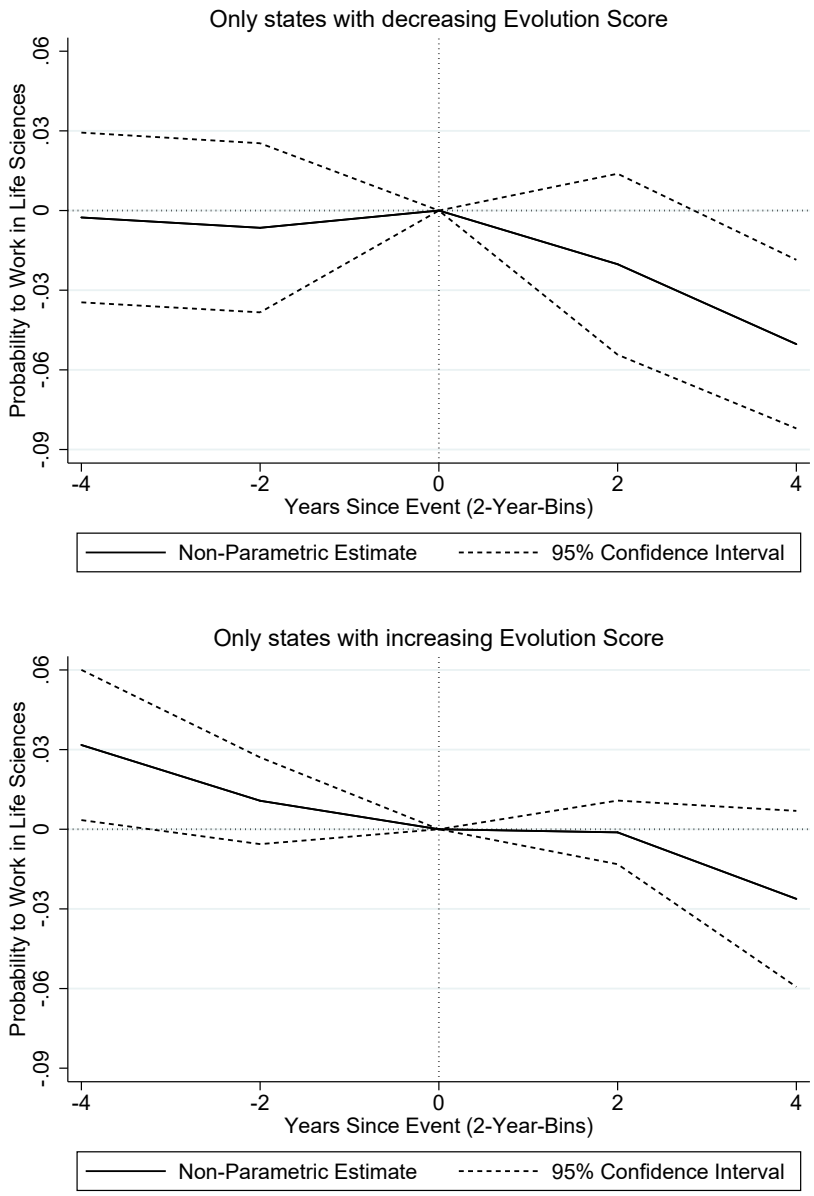
Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Share of questions about evolution answered correctly. Controls: Indicator variables for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate indicator variables for computer and books), as well as state, birth year and test year fixed effects. Numbers on horizontal axis refer to final year of respective two-year bins; i.e., 0 = last two years prior to treatment (excluded category), 2 = first two years of treatment. Inference: Clustering at state level. For the group of states with decreasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.183 and 0.009, respectively. For the group of states with increasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.609 and 0.181, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, 1996-2009 National Assessment of Educational Progress

Figure 3.6: Event-study graphs: Evolution approval in adulthood



Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know). Controls: Indicator variables for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (indicator variables for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Numbers on horizontal axis refer to final year of respective two-year bins; i.e., 0 = last two years prior to treatment (excluded category), 2 = first two years of treatment. Inference: Clustering at state level. For the group of states with decreasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.804 and 0.028, respectively. For the group of states with increasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.228 and 0.167, respectively. Data source: General Social Survey.

Figure 3.7: Event-study graphs: Occupational choice



Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). Controls: Indicator variables for gender, races/ethnicities, and survey year fixed effects. Numbers on horizontal axis refer to final year of respective two-year bins; i.e., 0 = last two years prior to treatment (excluded category), 2 = first two years of treatment. Inference: Clustering at state level. For the group of states with decreasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.901 and 0.007, respectively. For the group of states with increasing evolution score, the p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.087 and 0.247, respectively. Data source: General Social Survey.

Appendix

A3.1 Data Appendix

NAEP: Evolution Knowledge in School

The NAEP is a congressionally mandated project also known as the Nation's Report Card. It is administered by the National Center for Education Statistics (NCES), a body within the Institute of Education Sciences (IES) and the US Department of Education. Throughout the paper I use data from the Main-NAEP and not the Long-Term Trend NAEP, as the Main-NAEP has much larger sample sizes, is state-representative and, particularly relevant for this analysis, also covers science.

I categorize a question as addressing evolution if it contains the words "evolution" or "natural selection", or if it contains words that are based on the same word stem, such as "evolutionary".[1] I transform each question into a binary variable that is set equal to one if the correct answer was given, and equal to zero for any other answer, whether it is incorrect, partially correct, off task, etc. (the specific available categories depend on the question type). Figure A3.3 presents two sample questions, one on general Darwinian theory, and one on evolutionary trees. For each student, I calculate the share of questions on evolution that the student answered correctly. This share serves as the main outcome variable measuring a student's knowledge on evolution. I analogously group questions into nine categories of scientific topics other than evolution.[2] Table A3.12 shows that knowledge on evolution is positively correlated with knowledge on any of the other nine scientific topics.

In the preferred sample cut of keeping individuals who enter high school after 1990 and before 2010, I use the NAEP tests for science in grade 12 from 1996, 2000, 2005, and 2009. Regarding missings, I keep all students without missings on basic controls such as gender, and who come from birth cohorts of at least 10 observations. I set missings of other control variables to zero and add separate explanatory binary variables to account for these missings.[3]

The descriptive statistics for the main treatment, outcome, and control variables are presented in Table A3.13. The treatment variable "evolution score" captures the score of the evolution coverage of the State Science Standard in power in the state and year of a student's high

[1] Sometimes, the dataset does not contain the full wording of the questions but question keywords due to data protection reasons. I code such cases analogously, i.e. as addressing evolution if their keywords contain the words "evolution" or "natural selection", or if they contain words that are based on the same word stem.

[2] Notably, the number of questions available for each scientific topic in the pool of NAEP questions differs across scientific topics. Furthermore, each student receives only a subset of the pool of questions during the test. This test design explains why the number of questions answered on a given scientific topic differs across students. To address this issue, I calculate the share of questions answered correctly on a given scientific topic instead of the number of questions answered correctly. Moreover, this test design also explains why the number of students answering questions on a given scientific topic differs across scientific topics, resulting in varying sample sizes across scientific topics. These sample size differences are not a result of spurious selection, but are induced by the test design.

[3] The results are robust to not imputing the missings, as shown in Table A3.6.

school entry. The average evolution score equals 0.65, implying that students were on average exposed to a "satisfactory" evolution coverage.[4] The main outcome variable "evolution knowledge" is defined as the share of questions on evolution a students answers correctly. The fact that only 32 percent of questions on evolution are answered correctly on average underscores the difficulty of the test. For instance, the shares of students giving correct answers to the sample questions reported in Figure A3.3 equal 54 percent and 28 percent, respectively. Regarding non-evolution scientific topics, the average share of questions answered correctly amounts to 35 percent, indicating that the average difficulty of questions on evolution is largely similar to the overall difficulty. With regards to control variables, about half of the sample are female (51 percent). The shares of Whites, Blacks, Hispanics and Asians amount to 57 percent, 19 percent, 16 percent, and 6 percent, respectively. The various variables on the socioeconomic status indicate that a non-negligible share of students from grade 12 lives in underprivileged circumstances as measured by subsidized lunch status (30 percent), having no PC at home (16 percent), or disability status (11 percent).

GSS: Evolution Approval in Adulthood

The GSS data in the main sample comes from the waves from 2006, 2010, 2012, 2014, and 2016. Regarding correlations, the approval of evolution is almost only positively correlated with the other scientific outcomes, see Table A3.14. For all religious variables, I find a negative raw correlation with evolution approval as is visible in Table A3.15. The correlations between the political variables and evolution approval depend on the specific variable, see Table A3.16. For example, being in favor of same-sex marriage or marijuana legalization is positively correlated with evolution approval, while there is a negative correlation for identifying as Republican or being in favor of prayers in public schools.

Table A3.17 shows the descriptive statistics for the main treatment, outcome and control variables. The individuals in the sample were exposed to an evolution score of 0.63 on average which is very similar to corresponding sample average in NAEP, as expected given the comparable sample cut. Regarding the main outcome variable evolution approval, I find that 58 percent of sample say that the aforementioned statement about evolution is true. Regarding non-evolution scientific topics, six of the nine non-evolution scientific topics display higher rates of correct answers than evolution, with an average of 64 percent across these nine topics. Looking at religious outcomes, I note that 87 percent of respondents believe in God, and 70 percent are affiliated with a church. To give examples on political and partisan topics, 94 percent come out in favor of sex education in public schools, while the approval rates of same-sex marriage (64 percent) and abortion (46 percent) are considerably lower in this sample. With regard to the religious upbringing of these individuals, I observe that the most common religion/denomination an individual was raised in is Mainline Protestantism (37 percent), followed by Catholicism (32 percent), Non-Religious/Agnosticism/Atheism (13 percent), and Evangelicalism (10 percent).

[4] Lerner (2000b) classifies evolution scores between 0.60 and 0.79 as "satisfactory".

ACS: Occupational Choice

The estimation sample combines ACS waves from 2000-2017. The descriptive statistics are presented in Table A3.18. For the treatment variable, I find that the average evolution score exposure equals 0.68, which is similar to the corresponding averages from the analyses using the NAEP and the GSS. Regarding the outcome variables, all indicator variables for occupational fields are multiplied by 100 to ease the readability of descriptive statistics and reform effects. Hence, the descriptive statistics including mean and standard deviation are multiplied by 100 as well. For example, the sample mean of respondents working in life sciences equals 0.15, which implies that 0.15 percent of the sample work in this field. 0.84 percent of the sample work in any scientific field. Out of all 26 occupational fields, the largest sample shares are found for respondents working in office (13.2 percent) and in sales (11.5 percent).

A3.2 Supplementary Tables and Figures

Table A3.1: Evolution scores and reform year, by state

State	Evolution Score: 2009	Evolution Score: 2000	Evolution Score Difference 2009 - 2000	Reform Year	Only One Reform Event
Alabama	0.21	0.09	0.12	2005	NO
Alaska	0.59	0.48	0.11	2006	NO
Arkansas	0.66	0.55	0.11	2005	YES
DC	0.96	0.80	0.16	2006	YES
Florida	0.91	0.16	0.75	2008	YES
Georgia	0.66	0.07	0.59	2004	YES
Illinois	0.82	0.45	0.37	2004	YES
Kansas	0.96	0.00	0.96	2007	NO
Maine	0.68	0.30	0.38	2007	YES
Massachusetts	0.84	0.82	0.02	2006	NO
Minnesota	0.89	0.86	0.03	2009	NO
Mississippi	0.86	0.05	0.81	2008	NO
Nevada	0.77	0.70	0.07	2004	YES
New Hampshire	0.91	0.23	0.68	2006	YES
New Mexico	0.91	0.73	0.18	2003	YES
North Dakota	0.64	0.09	0.55	2006	NO
Ohio	0.86	0.28	0.58	2006	NO
Pennsylvania	0.96	0.91	0.05	2002	YES
Tennessee	0.55	0.02	0.53	2007	NO
Virginia	0.68	0.50	0.18	2003	YES
West Virginia	0.46	0.03	0.43	2008	NO
Wyoming	0.61	0.36	0.25	2003	YES
Colorado	0.82	0.86	-0.04	2009	NO
Connecticut	0.59	1.00	-0.41	2004	YES
Delaware	0.80	0.91	-0.11	2006	YES
Hawaii	0.75	0.91	-0.16	2005	YES
Indiana	0.96	1.00	-0.04	2006	NO
Louisiana	0.27	0.64	-0.37	2005	NO
Maryland	0.73	0.77	-0.04	2002	NO
Michigan	0.80	0.84	-0.04	2000	YES
Missouri	0.78	0.82	-0.04	2008	NO
Montana	0.75	0.82	-0.07	2006	YES
North Carolina	0.82	1.00	-0.18	2004	YES
Rhode Island	0.82	1.00	-0.18	2006	YES
South Carolina	0.91	0.95	-0.04	2005	NO
South Dakota	0.77	0.82	-0.05	2005	YES
Texas	0.46	0.64	-0.18	2009	YES
Arizona	0.82	0.82	0.00	-	-
California	1.00	1.00	0.00	-	-
Idaho	0.82	0.82	0.00	-	-
Iowa	0.77	No Standard	-	-	-
Kentucky	0.55	0.55	0.00	-	-
Nebraska	0.66	0.66	0.00	-	-
New Jersey	1.00	1.00	0.00	-	-
New York	0.68	0.68	0.00	-	-
Oklahoma	0.25	0.25	0.00	-	-
Oregon	0.82	0.82	0.00	-	-
Utah	0.82	0.82	0.00	-	-
Vermont	0.86	0.86	0.00	-	-
Washington	0.86	0.86	0.00	-	-
Wisconsin	0.55	0.55	0.00	-	-

Notes: Table reports the evolution score from 2009 based on Mead and Mates (2009), the evolution score from 2000 based on Lerner (2000b), and the difference of the evolution scores (evolution score from 2009 minus evolution score from 2000). States are listed in two pannels, postive or negative/zero evolution score change. For states that changed their evolution score, the respective year of the (last) reform as noted in Mead and Mates (2009) is also provided, and whether this reform is the only reform event between 2000 and 2009. The latter information on the only reform event is based on Gross (2005), Swanson (2005) as well as my own examination of state education websites.

Table A3.2: Effect of evolution coverage in Science Standards on probability to work in life sciences, by subgroups

	By Gender		By Race/Ethnicity		
	(1) Females	(2) Males	(3) Whites	(4) Blacks	(5) Hispanics
Evolution Score	0.052*	0.018	0.038*	0.012	0.004
	(0.020)	(0.020)	(0.016)	(0.016)	(0.034)
State FEs	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.15	0.14	0.16	0.04	0.06
Std. Dev. of Dep. Var.	3.92	3.75	4.05	2.06	2.44
Adj. R-squared	0.00068	0.00063	0.00047	0.00022	0.00030
Observations	3,220,042	3,240,608	5,023,449	789,587	765,295

Notes: Regressions by selected subgroups, as indicated in the columns headers. Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). Controls: Dummies for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table A3.3: Effect of evolution coverage in Science Standards on evolution knowledge in school: CS estimator

	Evolution Knowledge	
	States with decreasing evolution score	States with increasing evolution score
	(1)	(2)
Evolution Score	0.056***	0.028
	(0.017)	(0.022)

Notes: Dependent variable: Share of questions about evolution answered correctly. CS estimator (Callaway and Sant'Anna, 2021), accounting for heterogeneous treatment effects and staggered treatment timing. Simple aggregation of absolute value of all post treatment effects, using doubly robust inverse probability weighting. Controls: Never treated observations and not yet treated observations. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table A3.4: Effect of evolution coverage in Science Standards on evolution approval in adulthood: CS estimator

	Evolution Approval	
	States with decreasing evolution score	States with increasing evolution score
	(1)	(2)
Evolution Score	0.274	0.198
	(0.194)	(0.156)

Notes: Dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know). CS estimator (Callaway and Sant'Anna, 2021), accounting for heterogeneous treatment effects and staggered treatment timing. Simple aggregation of absolute value of all post treatment effects, using doubly robust inverse probability weighting. Controls: Never treated observations and not yet treated observations. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey

Table A3.5: Effect of evolution coverage in Science Standards on probability to work in life sciences: CS estimator

	Life Sciences	
	States with decreasing evolution score	States with increasing evolution score
	(1)	(2)
Evolution Score	0.036*	0.007
	(0.016)	(0.011)

Notes: Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). CS estimator (Callaway and Sant'Anna, 2021), accounting for heterogeneous treatment effects and staggered treatment timing. Simple aggregation of absolute value of all post treatment effects, using doubly robust inverse probability weighting. Controls: Never treated observations and not yet treated observations. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey

Table A3.6: Effect of evolution coverage in Science Standards on evolution knowledge, further robustness checks

	(1) Close Elections	(2) Control: Governor's Party	(3) State Specific Time Trends	(4) Only One Reform Event	(5) Sample Start: 1995	(6) Sample Start: 2000	(7) Outcome Coding: Indicator Variation	(8) Logit	(9) Probit	(10) Drop Missings
Evolution Score	0.083***	0.059**	0.042	0.079**	0.042*	0.033	0.051*	0.065*	0.063*	0.063**
	(0.021)	(0.020)	(0.068)	(0.022)	(0.016)	(0.023)	(0.022)	(0.032)	(0.031)	(0.022)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.33	0.32	0.32	0.32	0.31	0.30	0.35	0.32	0.32	0.33
Std. Dev. of Dep. Var.	0.42	0.42	0.42	0.42	0.42	0.44	0.43	0.42	0.42	0.42
Adj. R-squared	0.046	0.049	0.048	0.049	0.045	0.038	0.041	0.083	0.083	0.043
Observations	5,200	15,520	15,520	7,000	14,030	11,390	14,470	15,510	15,510	13,550

Notes: Robustness checks indicated in the column headers as follows: (1) Sample only includes states where the members of the State Board of Education are appointed by the governor, and where the governor in office at the time of the reform was voted into office with a margin of less than 10 percentage points compared to the runner-up; (2) Regressions control for political affiliation of governor ruling in the state and year of the student's high school entry; (3) Regressions include state-specific linear and quadratic time trends; (4) Sample only includes individuals from states that had only one reform event between 2000 and 2009, see Table A3.1 for more details; (5) Sample only includes individuals who entered high school after 1994; (6) Sample only includes individuals who entered high school after 1999; (7) Re-coding of dependent variable: Share of questions about evolution answered correctly. Indicator variable, 1=true, 0=false, missing=omitted/not reached/off-task/etc. (dependent on the question type); (8) Re-coding of dependent variable: Individuals are dropped from the sample if the question on evolution approval contains the word "elephants" instead of "human beings" (9) Coefficient reports average marginal treatment effect of logit specification; (10) Coefficient reports average marginal treatment effect of probit specification; (11) Sample excludes individuals who have missing values on controls instead of imputing them. Dependent variable: Share of questions about evolution answered correctly. Controls: Dummies for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate dummies for computer and books), and test year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table A3.7: Effect of evolution coverage in Science Standards on evolution approval in adulthood, further robustness checks

	(1) Close Elections	(2) Control: Governor's Party	(3) State Specific Time Trends	(4) Only One Reform event	(5) Sample Start: 1995	(6) Sample Start: 2000	(7) Outcome Coding Variation 1	(8) Outcome Coding Variation 2	(9) Logit	(10) Probit	(11) Drop Missings
Evolution Score	0.605**	0.332**	0.625**	0.394*	0.257*	0.313	0.288*	0.426**	0.347**	0.329*	0.304*
	(0.188)	(0.112)	(0.218)	(0.163)	(0.116)	(0.171)	(0.138)	(0.145)	(0.129)	(0.130)	(0.136)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.58	0.58	0.58	0.57	0.59	0.61	0.66	0.55	0.58	0.58	0.58
Std. Dev. of Dep. Var.	0.49	0.49	0.49	0.49	0.49	0.49	0.47	0.50	0.49	0.49	0.49
Adj. R-squared	0.102	0.107	0.096	0.115	0.092	0.077	0.127	0.102	0.117	0.117	0.110
Observations	589	1,801	1,801	709	1,299	654	1,571	1,617	1,780	1,780	1,751

Notes: Robustness checks indicated in the column headers as follows: (1) Sample only includes states where the members of the State Board of Education are appointed by the governor, and where the governor in office at the time of the reform was voted into office with a margin of less than 10 percentage points compared to the runner-up; (2) Regressions control for political affiliation of governor ruling in the state and year of the student's high school entry; (3) Regressions include state-specific linear and quadratic time trends; (4) Sample only includes individuals who entered high school after 1994; (5) Sample only includes individuals from states that had only one reform event between 2000 and 2009, see Table A3.1 for more details; (6) Sample only includes individuals who entered high school after 1999; (7) Sample excludes individuals whose dependent variable question on evolution replaces the words "human beings" with the word "elephants"; (8) Recoding of dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; missing=don't know); (9) Coefficient reports average marginal treatment effect of logit specification; (10) Coefficient reports average marginal treatment effect of probit specification; (11) Sample excludes individuals who have missing values on parental controls (parents born abroad, parental education) instead of imputing these values. Dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know; unless noted otherwise). Controls: Dummies for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (dummies for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table A3.8: Effect of evolution coverage in Science Standards on probability to work in life sciences, further robustness checks

	(1) Close Elections	(2) Control: Governor's Party	(3) State Specific Time Trends	(4) Only One Reform Event	(5) Sample Start: 1995	(6) Sample Start: 2000	(7) Logit	(8) Probit
Evolution Score	0.039	0.036*	0.025	0.031	0.036**	0.029*	0.033	0.035
	(0.025)	(0.014)	(0.025)	(0.021)	(0.013)	(0.012)	(0.028)	(0.026)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.14	0.15	0.15	0.14	0.13	0.10	0.15	0.15
Std. Dev. of Dep. Var.	3.80	3.84	3.84	3.77	3.58	3.10	3.84	3.84
Adj. R-squared	0.001	0.001	0.001	0.001	0.001	0.001		
Pseudo R-squared							0.032	0.032
Observations	2,022,927	6,460,650	6,460,650	2,522,283	4,821,487	2,762,694	6,460,650	6,460,650

Notes: Robustness checks indicated in the column headers as follows: (1) Sample only includes states where the members of the State Board of Education are appointed by the governor, and where the governor in office at the time of the reform was voted into office with a margin of less than 10 percentage points compared to the runner-up; (2) Regressions control for political affiliation of governor ruling in the state and year of the student's high school entry; (3) Regressions include state-specific linear and quadratic time trends; (4) Sample only includes individuals from states that had only one reform event between 2000 and 2009, see Table A3.1 for more details; (5) Sample only includes individuals who entered high school after 1994; (6) Sample only includes individuals who entered high school after 1999; (7) Coefficient reports average marginal treatment effect of logit specification; (8) Coefficient reports average marginal treatment effect of probit specification. Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). Controls: Dummies for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table A3.9: Effect of evolution coverage in Science Standards on evolution knowledge in school, by evolution score indicator variables

	Evolution Knowledge								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Evolution Score > 0.90	0.028								
	(0.018)								
Evolution Score > 0.80		0.013							
		(0.014)							
Evolution Score > 0.70			0.018						
			(0.013)						
Evolution Score > 0.60				0.023*					
				(0.009)					
Evolution Score > 0.50					0.023*				
					(0.009)				
Evolution Score > 0.40						0.018			
						(0.011)			
Evolution Score > 0.30							0.025*		
							(0.012)		
Evolution Score > 0.20								0.032**	
								(0.010)	
Evolution Score > 0.10									0.032**
									(0.010)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Std. Dev. of Dep. Var.	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
Adj. R-squared	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049
Observations	15520	15520	15520	15520	15520	15520	15520	15520	15520

Notes: Dependent variable: Share of questions about evolution answered correctly. Explanatory variables: Evolution score indicator variables (equals one if evolution score is larger than indicated level, and zero otherwise). Controls: Dummies for gender, races/ethnicities, subsidized lunch status, English language learner status, disability status, parental education, home possessions (separate dummies for computer and books), and test year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table A3.10: Effect of evolution coverage in Science Standards on evolution approval in adulthood, by evolution score indicator variables

	Evolution Knowledge								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Evolution Score > 0.90	0.115								
	(0.114)								
Evolution Score > 0.80		0.174*							
		(0.073)							
Evolution Score > 0.70			0.199*						
			(0.092)						
Evolution Score > 0.60				0.125					
				(0.068)					
Evolution Score > 0.50					0.138				
					(0.072)				
Evolution Score > 0.40						0.245***			
						(0.060)			
Evolution Score > 0.30							0.219**		
							(0.071)		
Evolution Score > 0.20								0.150	
								(0.109)	
Evolution Score > 0.10									0.071
									(0.107)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
Std. Dev. of Dep. Var.	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Adj. R-squared	0.102	0.104	0.104	0.104	0.104	0.106	0.104	0.103	0.102
Observations	1,801	1,801	1,801	1,801	1,801	1,801	1,801	1,801	1,801

Notes: Dependent variable: Approval to Evolution ("Human beings, as we know them today, developed from earlier species of animals - Is that true or false?", Indicator variable, 1=true, 0=false; don't know). Controls: Dummies for gender, races/ethnicities, parents born abroad, parental education, having lived with parents in adolescence, raised in rural area, religion raised in (dummies for mainline protestantism, evangelical protestantism, catholicism, no religion, judaism, buddhism, hinduism, other eastern, islam, orthodox-christian, christian, native american, inter-nondenominational, other religion), and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table A3.11: Effect of evolution coverage in Science Standards on probability to work in life sciences, by evolution score indicator variables

	Life Sciences								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Evolution Score > 0.90	0.003								
	(0.019)								
Evolution Score > 0.80		0.022*							
		(0.010)							
Evolution Score > 0.70			0.023*						
			(0.009)						
Evolution Score > 0.60				0.012					
				(0.007)					
Evolution Score > 0.50					0.013				
					(0.008)				
Evolution Score > 0.40						0.019			
						(0.010)			
Evolution Score > 0.30							0.020		
							(0.012)		
Evolution Score > 0.20								0.036**	
								(0.012)	
Evolution Score > 0.10									0.027**
									(0.010)
State FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Birth Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of Dep. Var.	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Std. Dev. of Dep. Var.	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84
Adj. R-squared	0.00064	0.00064	0.00064	0.00064	0.00064	0.00064	0.00064	0.00064	0.00064
Observations	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650	6,460,650

Notes: Dependent variable: Probability to work in life sciences (multiplied by 100 for interpretability). Explanatory variables: Evolution score dummies (equals one if evolution score is larger than indicated level, and zero otherwise). Controls: Dummies for gender, races/ethnicities, and survey year fixed effects. Standard errors clustered at the state level in parenthesis. Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: American Community Survey.

Table A3.12: Correlation coefficients of knowledge about different scientific areas

	Evolution	Motion	Matter and Mass	Energy	Reproduction	Climate	Pollution	Earth	Tectonics	Universe
Evolution	1									
Motion	0.0894***	1								
Matter and Mass	0.0836***	0.0115	1							
Energy	0.129***	0.0867***	0.250***	1						
Reproduction	0.283***	0.0132	0.156***	0.118***	1					
Climate	0.0524***	0.0694***	0.310***	0.183***	0.116***	1				
Pollution	0.150***	0.171***	0.0833***	0.0874***	0.139***	0.561***	1			
Earth	0.0924***	0.139***	-0.0451***	-0.0234**	0.122***	0.0632***	0.201***	1		
Tectonics	0.0183	0.0306	0.208***	0.228***	-0.0242	0.222***	0.174***	-0.00917	1	
Universe	0.117***	0.0435*	0.229***	0.115***	0.0544***	0.167***	0.0910	0.103***	0.188***	1

Notes: Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table A3.13: Descriptive statistics of NAEP data

	Mean	Std. Dev.	Min.	Max.
Treatment Variable:				
Evolution Score	0.65	0.31	0.00	1.00
Main Outcome:				
Evolution Knowledge	0.32	0.42	0.00	1.00
Placebo Outcomes - Non-Evolution Scientific Topics:				
Motion	0.51	0.43	0.00	1.00
Matter and Mass	0.30	0.43	0.00	1.00
Energy	0.38	0.43	0.00	1.00
Reproduction	0.38	0.42	0.00	1.00
Climate	0.39	0.39	0.00	1.00
Pollution	0.15	0.28	0.00	1.00
Earth	0.41	0.42	0.00	1.00
Tectonics	0.17	0.27	0.00	1.00
Universe and Big Bang	0.32	0.42	0.00	1.00
Non-Evolution Scientific Topics: Average	0.35	0.28	0.00	1.00
Controls:				
Female	0.51	0.50	0.00	1.00
Race/Ethnicity: White	0.57	0.49	0.00	1.00
Race/Ethnicity: Black	0.19	0.39	0.00	1.00
Race/Ethnicity: Hispanic	0.16	0.37	0.00	1.00
Race/Ethnicity: Asian	0.06	0.23	0.00	1.00
Race/Ethnicity: Other	0.01	0.11	0.00	1.00
English Language Learner	0.05	0.22	0.00	1.00
Disabled	0.11	0.32	0.00	1.00
Subsidized Lunch	0.30	0.46	0.00	1.00
Parental Education: Did not finish High School	0.09	0.29	0.00	1.00
Parental Education: Graduated High School	0.20	0.40	0.00	1.00
Parental Education: Some education after High School	0.26	0.44	0.00	1.00
Parental Education: Graduated College	0.44	0.50	0.00	1.00
Computer at Home	0.84	0.37	0.00	1.00
Books at Home: 0−10	0.23	0.42	0.00	1.00
Books at Home: 11−25	0.27	0.44	0.00	1.00
Books at Home: 26−100	0.33	0.47	0.00	1.00
Books at Home: >100	0.17	0.38	0.00	1.00

Notes: Descriptive statistics (mean, standard deviation, minimum, maximum) for treatment, outcome, and control variables. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12

Table A3.14: Correlation coefficients of approval to different scientific areas

	Evolution	Earth	Radioactivity	Reproduction	Lasers	Electrons	Viruses	Big Bang	Tectonics	Sun
Evolution	1									
Earth	0.120***	1								
Radioactivity	0.145***	0.196***	1							
Reproduction	-0.0222	0.0520*	0.0967***	1						
Lasers	0.106***	0.170***	0.233***	0.0323	1					
Electrons	0.169***	0.118***	0.177***	0.0883***	0.187***	1				
Viruses	0.107***	0.0925***	0.247***	0.183***	0.172***	0.170***	1			
Big Bang	0.415***	0.167***	0.139***	0.0540*	0.130***	0.141***	0.109***	1		
Tectonics	0.248***	0.211***	0.153***	0.0820***	0.107***	0.156***	0.119***	0.240***	1	
Sun	0.109***	0.157***	0.189***	0.0922***	0.171***	0.133***	0.216***	0.179***	0.187***	1

Notes: Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table A3.15: Correlation coefficients of evolution approval and religious outcomes

	Evolution Approval	God	Bible	Afterlife	Rebirth	Strong Believer	Religious Affiliation	Church- going	Church Activities	Personal Prayer	Missionize	Spiritual Person	Religious Person	Fundamentalist	Religious Outcomes: Average
Evolution Approval	1														
God	-0.194***	1													
Bible	-0.272***	0.420***	1												
Afterlife	-0.106***	0.378***	0.368***	1											
Rebirth	-0.313***	0.222***	0.339***	0.256***	1										
Strong Believer	-0.284***	0.230***	0.287***	0.242***	0.412***	1									
Religious Affiliation	-0.212***	0.374***	0.450***	0.326***	0.289***	0.452***	1								
Church-going	-0.267***	0.221***	0.324***	0.227***	0.415***	0.510***	0.390***	1							
Church Activities	-0.203***	0.150***	0.208***	0.159***	0.343***	0.404***	0.218***	0.527***	1						
Personal Prayer	-0.282***	0.423***	0.453***	0.343***	0.378***	0.392***	0.455***	0.417***	0.289***	1					
Missionize	-0.275***	0.242***	0.348***	0.259***	0.535***	0.416***	0.294***	0.417***	0.325***	0.408***	1				
Spiritual Person	-0.158***	0.292***	0.210***	0.292***	0.320***	0.385***	0.280***	0.319***	0.267***	0.422***	0.339**	1			
Religious Person	-0.241***	0.290***	0.382***	0.298***	0.395***	0.547***	0.434***	0.465***	0.326***	0.474***	0.438**	0.494***	1		
Fundamentalist	-0.248***	0.162***	0.249***	0.157***	0.321***	0.284***	0.378***	0.276***	0.223***	0.258***	0.337**	0.181***	0.295***	1	
Religious Outcomes: Average	-0.374***	0.519***	0.615***	0.526***	0.650***	0.693***	0.658***	0.683***	0.537***	0.707***	0.669**	0.605***	0.731***	0.500***	1

Notes: Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table A3.16: Correlation coefficients of evolution approval and political outcomes

	Evolution Approval	Republican	Conservative	Prayer in Public Schools	Sex Education in Public Schools	Same-Sex Marriage	Abortion	Marijuana Legalization	Capital Punishment	Gun Control	Immigration	Environment	Alternative Energy Sources	Education	Scientific Research	Reducing Income Differences	Assistance to the Poor	Conditions of Blacks
Evolution Approval	1																	
Republican	-0.120***	1																
Conservative	-0.126***	0.430***	1															
Prayer in Public Schools	-0.236***	0.123***	0.190***	1														
Sex Education in Public Schools	0.198***	-0.0833**	-0.156***	-0.0545	1													
Same-Sex Marriage	0.287***	-0.158***	-0.241***	-0.253***	0.262***	1												
Abortion	0.240***	-0.161***	-0.197***	-0.225***	0.108*	0.278***	1											
Marijuana Legalization	0.128***	-0.126***	-0.159***	-0.163***	0.124***	0.184***	0.252***	1										
Capital Punishment	-0.0136	0.228***	0.146***	0.0630*	-0.00745	-0.0558	-0.0743*	-0.0167	1									
Gun Control	0.0279	-0.0962**	-0.0772*	-0.109*	0.0939*	0.106***	0.0219	-0.143***	-0.0486	1								
Immigration	0.00435	-0.110***	-0.0811**	-0.0407	-0.00363	0.135**	0.107*	0.0592	-0.121***	0.158***	1							
Environment	0.0890***	-0.150***	-0.170***	-0.0556	0.139***	0.164***	0.128***	0.0350	-0.0497*	0.0863**	0.0270	1						
Alternative Energy Sources	0.0880**	-0.0656*	-0.0862**	-0.124***	0.112**	0.193***	0.138***	0.175***	-0.0548	0.0275	0.0120	0.271***	1					
Education	0.0560*	-0.101***	-0.0700**	-0.0371	0.120***	0.0960**	0.140***	0.0993***	-0.0321	0.0273	-0.0144	0.192***	0.152***	1				
Scientific Research	0.163***	-0.0742**	-0.0120	-0.100***	0.0583*	0.120***	0.157***	0.0808**	-0.0164	0.00515	0.0431	0.138***	0.221***	0.0912***	1			
Reducing Income Differences	0.0912***	-0.219***	-0.176***	-0.00982	0.0789*	-0.00254	-0.0122	0.108***	-0.110***	0.0495	0.0289	0.160***	0.165***	0.119***	0.0478	1		
Assistance to the Poor	-0.00113	-0.156***	-0.128***	-0.0214	-0.000860	0.0445	0.0160	0.0950***	-0.119***	0.0756*	0.0416	0.122***	0.0272	0.121***	0.0202	0.190***	1	
Conditions of Blacks	0.0594*	-0.170***	-0.153***	-0.00191	0.0723*	0.0754*	0.0959**	0.0626*	-0.178***	0.0983**	0.114***	0.149***	0.0974***	0.147***	0.0963***	0.163***	0.185***	1

Notes: Single, double, and triple asterisks indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Data source: General Social Survey.

Table A3.17: Descriptive statistics of GSS data

	Mean	Std. Dev.	Min.	Max.
Treatment Variable:				
Evolution Score	0.63	0.33	0.00	1.00
Main Outcome:				
Evolution Approval	0.58	0.49	0.00	1.00
Placebo Outcomes - Non-Evolution Scientific Topics:				
Earth	0.88	0.33	0.00	1.00
Radioactivity	0.65	0.48	0.00	1.00
Reproduction	0.62	0.48	0.00	1.00
Lasers	0.47	0.50	0.00	1.00
Electrons	0.57	0.50	0.00	1.00
Viruses	0.51	0.50	0.00	1.00
Big Bang	0.45	0.50	0.00	1.00
Tectonics	0.85	0.36	0.00	1.00
Sun	0.79	0.41	0.00	1.00
Non-Evolution Scientific Topics: Average	0.64	0.22	0.00	1.00
Placebo Outcomes - Religious Attitudes:				
God	0.87	0.33	0.00	1.00
Bible	0.72	0.45	0.00	1.00
Afterlife	0.72	0.45	0.00	1.00
Rebirth	0.34	0.47	0.00	1.00
Strong Believer	0.32	0.47	0.00	1.00
Religious Affiliation	0.70	0.46	0.00	1.00
Church-going	0.35	0.48	0.00	1.00
Church Activities	0.17	0.38	0.00	1.00
Personal Prayer	0.65	0.48	0.00	1.00
Missionize	0.41	0.49	0.00	1.00
Spiritual Person	0.56	0.50	0.00	1.00
Religious Person	0.43	0.49	0.00	1.00
Fundamentalist	0.24	0.43	0.00	1.00
Religious Outcomes: Average	0.50	0.28	0.00	1.00
Placebo Outcomes - Political Attitudes:				
Republican	0.28	0.45	0.00	1.00
Conservative	0.26	0.44	0.00	1.00
Prayer in Public Schools	0.45	0.50	0.00	1.00
Sex Education in Public Schools	0.94	0.24	0.00	1.00
Same-Sex Marriage	0.64	0.48	0.00	1.00
Abortion	0.45	0.50	0.00	1.00
Marijuana Legalization	0.55	0.50	0.00	1.00
Capital Punishment	0.61	0.49	0.00	1.00
Gun Control	0.70	0.46	0.00	1.00
Immigration	0.15	0.36	0.00	1.00

Table A.17 (continued) – Descriptive statistics of GSS data

	Mean	Std. Dev.	Min.	Max.
Placebo Outcomes - Political Attitudes (continued):				
Environment	0.71	0.45	0.00	1.00
Alternative Energy Sources	0.59	0.49	0.00	1.00
Education	0.81	0.40	0.00	1.00
Scientific Research	0.37	0.48	0.00	1.00
Reducing Income Differences	0.53	0.50	0.00	1.00
Assistance to the Poor	0.49	0.50	0.00	1.00
Conditions of Blacks	0.36	0.48	0.00	1.00
Controls:				
Female	0.57	0.50	0.00	1.00
Race/Ethnicity: White	0.70	0.46	0.00	1.00
Race/Ethnicity: Black	0.19	0.39	0.00	1.00
Race/Ethnicity: Other	0.12	0.32	0.00	1.00
Race/Ethnicity: Hispanic	0.16	0.37	0.00	1.00
Raised in Rural Area	0.49	0.50	0.00	1.00
Parents born in US	0.19	0.39	0.00	1.00
Parents born abroad	0.81	0.39	0.00	1.00
Parental Education: No Highschool	0.11	0.31	0.00	1.00
Parental Education: Highschool	0.50	0.50	0.00	1.00
Parental Education: More than Highschool	0.39	0.49	0.00	1.00
Growing up: Both Parents	0.55	0.50	0.00	1.00
Growing up: One Parent, one Stepparent	0.12	0.33	0.00	1.00
Growing up: Single Parent	0.25	0.43	0.00	1.00
Growing up: Other	0.05	0.22	0.00	1.00
Raised as Protestant: Mainline	0.37	0.48	0.00	1.00
Raised as Protestant: Evangelical	0.09	0.29	0.00	1.00
Raised as Catholic	0.32	0.47	0.00	1.00
Raised as Jew	0.01	0.10	0.00	1.00
Raised as Non-Religious	0.14	0.34	0.00	1.00
Raised as Other	0.01	0.08	0.00	1.00
Raised as Buddhist	0.00	0.06	0.00	1.00
Raised as Hindu	0.00	0.05	0.00	1.00
Raised as Other Eastern Rel.	0.00	0.03	0.00	1.00
Raised as Muslim	0.00	0.06	0.00	1.00
Raised as Orthodox-Christian	0.00	0.05	0.00	1.00
Raised as Christian	0.04	0.20	0.00	1.00
Raised as Native American	0.00	0.03	0.00	1.00
Raised as Inter-Nondenominational	0.00	0.02	0.00	1.00

Notes: Descriptive statistics (mean, standard deviation, minimum, maximum) for treatment, outcome, and controls variables. Data source: General Social Survey.

Table A3.18: Descriptive statistics of ACS data

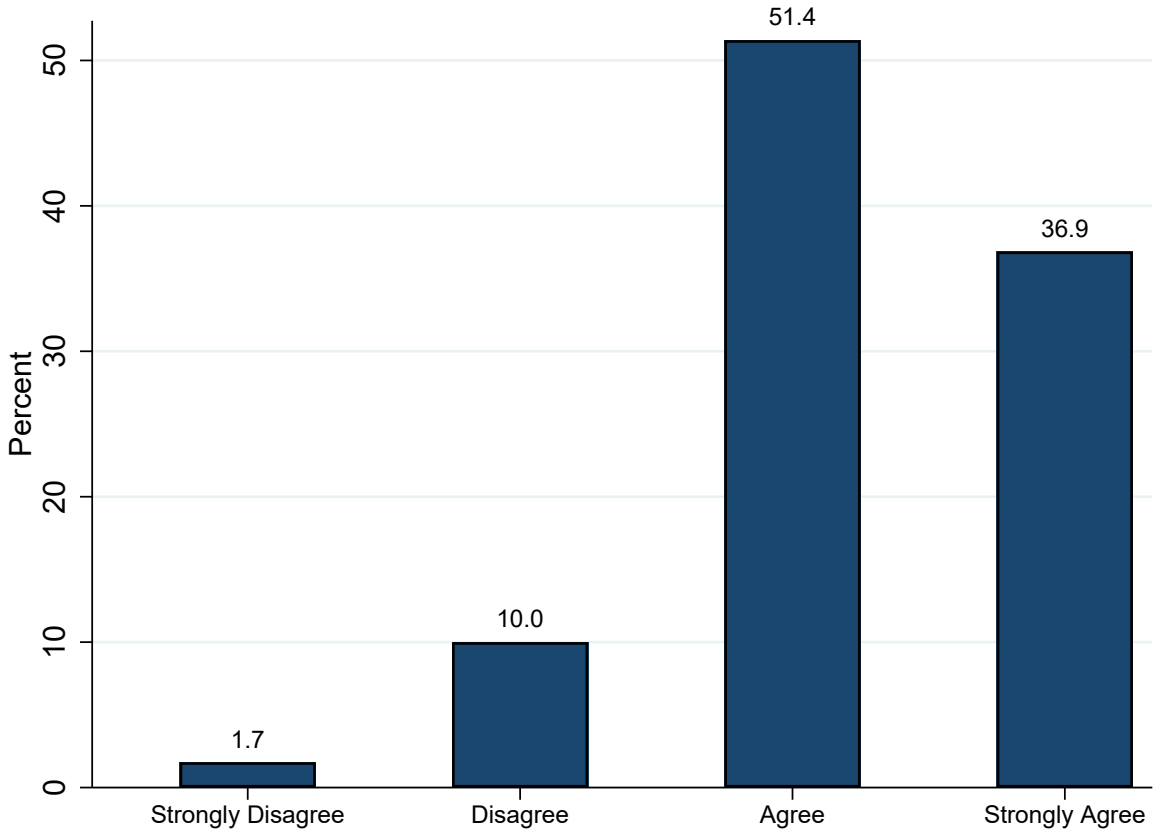
	Mean	Std. Dev.	Min.	Max.
Treatment Variable:				
Evolution Score	0.67	0.30	0.00	1.00
Main Outcomes - Working in scientific fields:				
Life Sciences	0.15	3.84	0.00	100.00
Physical Sciences	0.22	4.68	0.00	100.00
Social Sciences	0.16	4.03	0.00	100.00
Science Technicians	0.32	5.62	0.00	100.00
Overall: All Sciences	0.85	9.16	0.00	100.00
Placebo Outcomes - Working in non-scientific fields:				
Management	5.44	22.68	0.00	100.00
Analysts	1.87	13.54	0.00	100.00
Finance	1.63	12.67	0.00	100.00
IT	1.97	13.89	0.00	100.00
Engineering	1.39	11.69	0.00	100.00
Social	1.39	11.70	0.00	100.00
Legal	0.82	9.01	0.00	100.00
Education	5.58	22.96	0.00	100.00
Arts	2.04	14.15	0.00	100.00
Health Care	4.33	20.36	0.00	100.00
Health Care Support	2.63	15.99	0.00	100.00
Protective Services	2.13	14.43	0.00	100.00
Food	7.88	26.94	0.00	100.00
Buildings	2.84	16.61	0.00	100.00
Personal Care	3.96	19.51	0.00	100.00
Sales	11.42	31.81	0.00	100.00
Office	13.25	33.90	0.00	100.00
Farming	0.66	8.07	0.00	100.00
Construction	4.45	20.63	0.00	100.00
Extraction	0.21	4.60	0.00	100.00
Installation	2.80	16.50	0.00	100.00
Production	4.87	21.53	0.00	100.00
Transportation	5.35	22.50	0.00	100.00
Armed Forces	0.75	8.62	0.00	100.00
Unemployed / Not in Labor Market	9.48	29.30	0.00	100.00

Table A.18 (continued) – Descriptive statistics of ACS data

	Mean	Std. Dev.	Min.	Max.
Controls:				
Female	0.50	0.50	0.00	1.00
Race/Ethnicity: White	0.78	0.42	0.00	1.00
Race/Ethnicity: Black	0.12	0.33	0.00	1.00
Race/Ethnicity: Asian	0.02	0.16	0.00	1.00
Race/Ethnicity: Native	0.01	0.11	0.00	1.00
Race/Ethnicity: Other	0.03	0.18	0.00	1.00
Race/Ethnicity: Multiple	0.03	0.17	0.00	1.00
Race/Ethnicity: Hispanic	0.12	0.32	0.00	1.00

Notes: Descriptive statistics (mean, standard deviation, minimum, maximum) for treatment, outcome (multiplied by 100 for interpretability), and controls variables. Data source: American Community Survey.

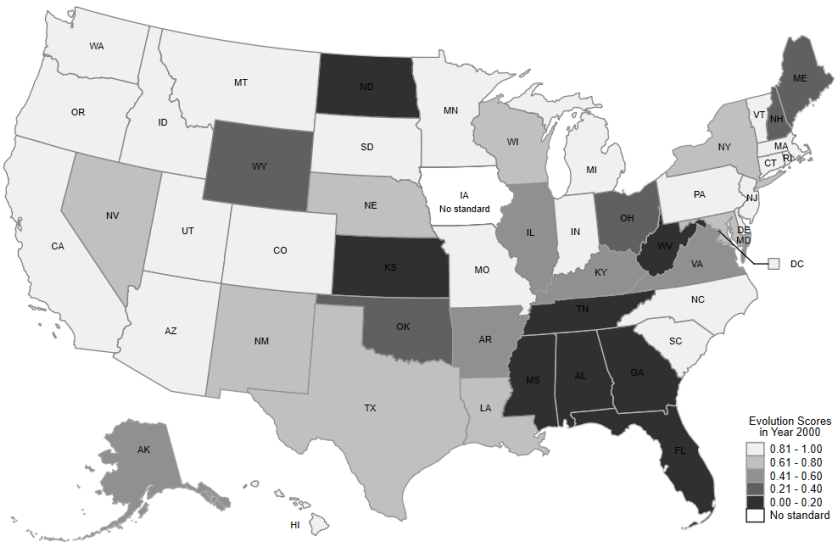
Figure A3.1: Teachers' focus on State Science Education Standards when teaching evolution (Teacher Survey)



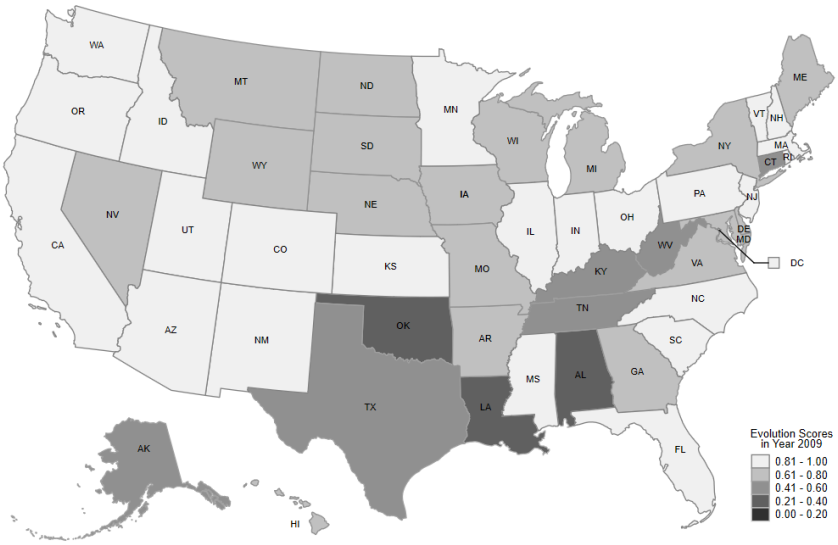
Note: Histogram depicts answer categories on agreement with the statement "When I do teach evolution, I focus heavily on what students need to know to meet state science standards.". Data is drawn from the National Survey for High School Biology Teachers conducted by the Survey Research Center (Penn State). Sample: Nationally representative of U.S. public high school biology teachers. Sample size: 926 teachers. Eligibility: Teachers in public school where grades 9 and 10 are offered, who taught a high school biology class in at least the previous year, and who had not recently retired. Survey date: March 1 to May 5, 2007. More information on dataset available in Berkman et al. (2008) and Berkman and Plutzer (2011). Data Source: National Survey for High School Biology Teachers

Figure A3.2: US map of evolution scores

Year 2000



Year 2009



Note: Map depicts the evolution score of US States in 2000 and 2009, respectively. The evolution score measures the coverage of evolution in Science Standards, as reported in Lerner (2000b) and Mead and Mates (2009). An evolution score of 0 indicates no or a non-scientific/creationist coverage of evolution, and a score of 1 a very comprehensive coverage of evolution. A list of the evolution scores underlying this map is provided in Table A3.1. Data source: Lerner (2000b) and Mead and Mates (2009)

Figure A3.3: Two NAEP sample questions on evolution knowledge

Sample Question 1

Which of the following is NOT a part of Darwin's theory of evolution by natural selection?

A. Individuals in a population vary in many ways.

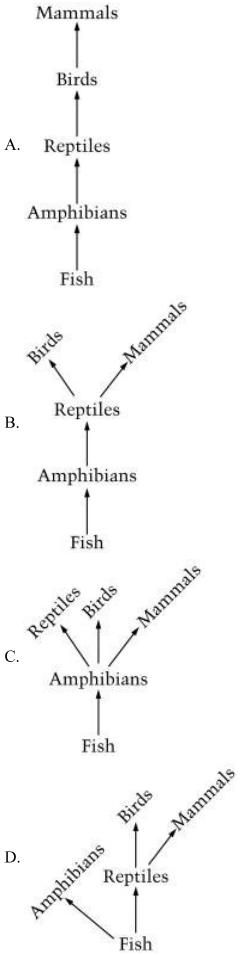
B. Some individuals possess features that enable them to survive better than individuals lacking those features.

C. More offspring are produced than can generally survive.

D. Changes in an individual's genetic material are usually harmful.

Sample Question 2

According to evolutionary theory, which of the following evolutionary trees best describes the relationship between groups of vertebrates?



Note: Sample question on evolution knowledge from NAEP Science Test, Grade 12, Year 2000. Question also accessible online at NAEP question tool. Question 1: Answer D is correct. Question 2: Answer B is correct. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment for Grade 12

4 The Unintended Effects of the Common Core State Standards on Non-Targeted Subjects*

4.1 Introduction

Student achievement in the United States has been lagging behind the student achievement of many other industrialized countries for a number of decades (Hanushek et al., 2012; Shakeel and Peterson, 2021). The adoption of rigorous centralized education standards as a means of aligning basic elements of local school curricula has long been proposed to raise US student achievement (Costrell, 1994; Bishop, 1997). In 2008, a report published by the National Governors Association titled "Benchmarking for Success" suggested that US states should adopt a common core of internationally benchmarked education standards (Jerald, 2008). Such a standard, named the "Common Core State Standards" (CCSS), was subsequently developed for math and English Language Arts (ELA). The CCSS did not include other subjects such as science and social studies. From 2010 onwards, states could voluntarily adopt the CCSS. By 2022, 42 states had adopted the CCSS (Achieve Inc., 2013; CCSSI, 2022).

The theoretical literature on the effects of centralizing education standards does not offer a clear prediction on whether adopting the CCSS increases student achievement.[1] The empirical literature on the effects of the CCSS on student achievement has so far documented zero to modest positive effects on student achievement in the targeted subjects math and ELA. We replicate this analysis in our setting and come to largely the same conclusion, although our results suggest that the prior literature rather overestimates than underestimates any positive effects on student achievement in targeted subjects.[2] However, the main focus of our paper is

* This chapter is joint work with M. Danish Shakeel. It is based on the paper 'The Unintended Effects of the Common Core State Standards on Non-Targeted Subjects', *Program on Education Policy and Governance Working Papers Series 21-03, Harvard University*, 2021.

[1] On the one hand, a centralized education standard could overcome the problem that education standards emerging from a decentralized process of setting standards tend to have inefficiently low degrees of rigor, potentially harming student achievement. This lack of rigor is caused by a free-riding problem induced by mobility of high school graduates across states and their pooling in the labor market (Costrell, 1994). This problem was of special relevance in the years before the adoption of the CCSS, as states had an incentive to adopt inefficiently low standards under the No Child Left Behind Act (NCLB) as means to increase pass rates of standardized tests (Maranto and Maranto, 2004; McCluskey and Coulson, 2007). On the other hand, a centralized education standard could also decrease student achievement by being less tailored to state-level preferences, and by abolishing the 'laboratory federalism' and competition between states for better education standards (Tiebout, 1956; Oates, 1999).

[2] Most of the early studies are correlational and find zero to modest positive associations between CCSS and student achievement in targeted subjects (Schmidt and Houang, 2012; Loveless, 2014, 2015, 2016). Exploiting quasi-experimental settings, more recent papers confirm the zero to modest positive effects (depending on

on spillovers of the CCSS on student achievement in non-targeted subjects. Such spillovers have, to the best of our knowledge, not yet been studied in a causal framework. We argue that studying such spillovers is essential for evaluating the overall success of the CCSS in terms of student achievement and for guiding future reforms of education standards in general.

To close this research gap, we estimate the effect of the CCSS on student achievement in non-targeted subjects such as science and social studies. In theory, it is unclear whether any spillovers of the CCSS on student achievement in non-targeted subjects are positive or negative (or zero). On the one hand, the CCSS could be beneficial beyond its target subjects if, for example, skills acquired in a targeted subject such as math help students to perform well in a non-targeted subject such as science. On the other hand, the CCSS could have caused a reduction of instructional focus on the non-targeted subjects, possibly leading to a decline in student achievement in those subjects.

Simple correlations between CCSS adoption and student achievement in non-targeted subjects likely do not yield a causal answer to our research question. States that adopted the CCSS plausibly differ from states that did not adopt the CCSS in ways that affect student achievement through many channels other than education standards, for example through differences in political preferences or human capital. To overcome this identification problem, we estimate the effect of the CCSS on student achievement in non-targeted subjects in a two-way fixed effects difference-in-differences (DD) framework. This approach builds on the idea that states without CCSS adoption in a certain year act as counterfactuals for states with CCSS adoption in that year, after accounting for time-invariant differences between states and national differences between years.

To run these DD models, we combine state-level data on the adoption of the CCSS with individual-level student achievement data from the National Assessment of Educational Progress (NAEP). The NAEP is known as The Nation's Report Card and provides unique student-level test score data for a large number of years, grades and subjects. In particular, it covers a range of subjects not targeted by the CCSS, namely science and different social studies (civics, economics, geography, and history). It is comparable across states and over time and covers the relevant years before and after the adoption of the CCSS. The NAEP student and teacher surveys complement the test score data by providing information on student characteristics, teacher characteristics and classroom instruction.

study and subgroups therein) for Kentucky (Xu and Cepa, 2018), California (Gao and Lafortune, 2019), Chicago (Allensworth et al., 2021) and a subset of US states (Bleiberg, 2021). An exception is Song et al. (2019) who find negative effects on student achievement in targeted subjects of the adoption of general College and Career Readiness Content Standards (CCRCS). This study cannot be directly compared to the other studies, as the CCRCS include the CCSS but also other education standards. In our paper, we also estimate the effect of the CCSS on student achievement in targeted subjects in our setting and, again, find zero to modest positive effects. This finding largely confirms the conclusions of the prior literature, although our robustness checks indicate that prior analyses rather overestimate any positive effects, see Appendix A4.2 for more details.

We find a significant negative effect of the CCSS on student achievement in non-targeted subjects. More specifically, being exposed to the CCSS for the entire school career (at the time of testing) as opposed to not being exposed to the CCSS at all decreases student achievement in non-targeted subjects on average by 0.08 units of a standard deviation. The effect size can be interpreted as a loss of learning worth approximately 25 percent to 30 percent of a school year. We regard this finding as reduced-form evidence that the CCSS induced a reduction of instructional focus on non-targeted subjects.

Next, we hypothesize that the negative effect is over-proportionally large for underprivileged students as these students (or their parents and environments in general) might be less able to compensate for the reduction of instructional focus. To test this hypothesis, we conduct subgroup analysis by student characteristics and find that the negative effect is mostly driven by Black and Hispanic students, and by students with free or reduced price lunch status, English language learner status, and disability status. We also conduct subgroup analyses by subject and grades and find that the achievement losses are most pronounced for science and for students in grade 4.[3] In sum, we conclude that the CCSS mainly reduced student achievement in non-targeted subjects among underprivileged students. This decline in student achievement exacerbates racial/ethnic and socioeconomic student achievement gaps as well as the achievement gap between students with and without disabilities in the non-targeted subjects.

A series of robustness checks supports our main results. We show that the results are robust to event-study specifications and specifications with state-specific time trends. To account for state-specific shocks simultaneous to the adoption of the CCSS, we run a triple-difference model with students from private schools as an additional control group for which the CCSS was never mandatory. We also control explicitly for a large list of educational reforms. Furthermore, we account for recent developments in the econometric literature on two-way-fixed effects models and time-varying treatment effects and show that our results are not driven by negative weights (Chaisemartin and D'Haultfœuille, 2020; Athey and Imbens, 2021; Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Roth and Sant'Anna, 2021; Sun and Abraham, 2021). Another set of robustness checks defines treatment based on information about actual CCSS implementation in the different states to account for the fact that CCSS adoption and CCSS implementation could diverge.

To uncover the mechanisms behind these results, we aim to understand what has changed in the classrooms of the students due to the CCSS. To this end, we draw on the NAEP teacher survey data on instructional focus. We find that the CCSS reduced teacher-reported instruction time, instructional resources, and some dimensions of the quality of teacher-student interactions for the non-targeted subjects. This finding suggests that the exclusion of science

[3] In addition to the CCSS, some states adopted the Next Generation Science Standards (NGSS). The NGSS were released in 2013 with slow take-up rates by states relative to the CCSS. In appendix A4.1, we demonstrate that our results are robust to controlling for NGSS adoption.

and social studies from the CCSS has signaled a lower relative importance of these subjects, resulting in a reduction of instructional focus.

This paper expands the literature on the existence and effects of strategic responses of educators to reforms of education systems. There is ample evidence that accountability reforms can shift the instructional focus away from subjects, topics, and students not targeted by the accountability rating (Jacob, 2005; Reback, 2008; Neal and Schanzenbach, 2010). The CCSS can be interpreted as an extension to the No Child Left Behind Act (NCLB) of 2002.[4] Analogous to our findings on the CCSS, NCLB reduced instruction time in non-targeted science and social studies (Reback et al., 2014). However, while we find for the CCSS that the reduction in instructional focus on the non-targeted subjects coincides with student achievement declines in these subjects, NCLB and other accountability reforms did not (necessarily) cause equivalent achievement declines (Winters et al., 2010; Dee and Jacob, 2011; Reback et al., 2014). One possible explanation for this difference is that the CCSS has led to much less pronounced student achievement gains in the targeted subjects relative to NCLB. This lack of substantial student achievement gains in the targeted subjects under CCSS may have limited positive spillovers on student achievement in non-targeted subjects, and may have made shifts in instructional focus matter more.

More generally, our paper contributes to the small but growing quasi-experimental literature on how the content of education standards affects individuals. Recently, it has been demonstrated that the content of US state education standards affects students' skills, attitudes, and occupational choice (Arold, 2022). Beyond education standards, the content of education in general influences skills (Cortes and Goodman, 2014; Goodman, 2019; Conger et al., 2021), labor market outcomes (Altonji et al., 2012; Fuchs-Schündeln and Masella, 2016) as well as identity, preferences and beliefs (Clots-Figueras and Masella, 2013; Cantoni et al., 2017; Bazzi et al., 2020).

Our outcome variable student achievement is not only interesting in its own right, but also an important predictor of economic outcomes at the individual and societal level. Student achievement has been found to affect earnings, income distribution, and economic growth (Hanushek and Woessmann, 2008, 2012). Notably, the predictive power of student achievement is much stronger than that of traditional measures of human capital used in the literature such as literacy rates (Romer, 1990), school enrollment (Barro, 1991), or years of education (Barro and Lee, 2013). Similarly, student achievement gaps between races/ethnicities (which we document for the non-targeted subjects) have been shown to account for relevant shares

[4] NCLB was an accountability reform that compelled states to design school accountability systems based on annual student assessments in math and ELA that were linked to state standards. To help states improve and meet the NCLB standards, Common Core provides education standards in math and ELA defining the knowledge and skills expected of students at each grade level. Furthermore, states that adopted the CCSS could get a waiver from some of the NCLB regulations. In this sense, our estimated effect of Common Core adoption combines both the impact of introducing Common Core standards and removing part of NCLB. Additional analyses presented in Section 4.6 suggest that the overall effect is not primarily coming from the NCLB waivers.

of the racial/ethnic gap in adulthood social and economic outcomes (Fryer, 2011). Although student achievement does not adequately capture non-cognitive skills (Heckman and Kautz, 2012; Jackson, 2018) which are increasingly important in the labor market (Deming, 2017), student achievement has been shown to be a strong predictor of not only cognitive skills, but also a broad set of individual-level outcomes including physical and mental health, and voting behavior (Borghans et al., 2016).

The paper proceeds as follows. Section 4.2 provides institutional background about the adoption and implementation of the CCSS. Section 4.3 outlines the empirical approach, while Section 4.4 describes the data. Section 4.5 reports the main results and heterogeneities of the effect of CCSS exposure on student achievement in non-targeted subjects. A series of robustness checks is presented in Section 4.6. Section 4.7 shows additional analyses of mechanisms, and Section 4.8 concludes.

4.2 Institutional Background

4.2.1 Background and Data on the Adoption of the CCSS

The idea of centralizing education standards in the US has been discussed for decades (Costrell, 1994, 1997; Betts, 1998). In 2008, the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), and Achieve Inc. jointly published a report titled "Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education." (Jerald, 2008). The report prescribed that US states adopt a common core of internationally benchmarked standards in math and ELA to raise US achievement levels on international assessments. A number of philanthropic organizations, including the Bill & Melinda Gates Foundation, provided resources to enable the states to establish a common core of standards. Subsequently, numerous state governments, teachers' unions and other interest groups advocated for a systemic change in education standards across the nation. In 2009, a consortium of the National Governors Association and the Council of Chief State School Officers, with support from the U.S. Department of Education, set incentives for states to adopt the CCSS. If a state adopted the CCSS, it could get a waiver from some of the No Child Left Behind (NCLB) regulations.[5]

Our primary source for state-level data on the adoption of the CCSS is Achieve Inc. (2013), with an updated version provided by CCSSI (2022). CCSSI (2022) is the website of the Common Core State Standards Initiative provided by the National Governors Association Center for Best Practices and the Council of Chief State School Officers. They report if and when a state has adopted the CCSS.[6] Based on this data source, 42 states have adopted the CCSS permanently. Of the states that have adopted the CCSS permanently, most states adopted it in 2010, while a

[5] NCLB was a federal legislation signed in 2002 by President Bush. In December 2015, President Obama signed the Every Student Succeeds Act (ESSA). ESSA replaced the NCLB act. ESSA shifted NCLB's federal accountability aspect to the states.

[6] Throughout the paper, we treat Washington D.C. as a state.

number of states adopted it in 2011 and one state, Wyoming, adopted it in 2012. In contrast, Alaska, Florida, Indiana, Nebraska, Oklahoma, South Carolina, Virginia and Texas have not adopted the CCSS permanently. In our baseline coding, we code students in these 8 states as never having been treated by the CCSS.[7] To account for the fact that some of those 8 states had adopted the CCSS temporarily, we present robustness checks in which we treat temporary adopters as treated from the year of the temporary adoption, even if they repealed/revised the CCSS later on. In the latter coding, only the states that never adopted the CCSS, Alaska, Nebraska, Virginia and Texas, remain in the control group, based on data from Bleiberg (2021) and CCSSI (2022). The map presented in Figure 4.1 illustrates the CCSS adoption graphically.

4.2.2 Background and Data on the Implementation of the CCSS

The implementation of the CCSS was not straightforward. There is anecdotal evidence that the CCSS presented challenges in teaching and testing to schools. Some teachers had difficulty adjusting to the new curriculum, and CCSS-based standardized tests were not always suitable. A case study in New York found that the CCSS led to exceedingly long and difficult exams. The rigor of the standardized tests exceeded the level of college readiness and represented more of an early college level (Polleck and Jeffery, 2017). In addition, some states did not have assessments and textbooks aligned with the CCSS until 2013 or later (Polikoff, 2017). Although not all challenges of the implementation of the CCSS have been overcome everywhere (Polikoff, 2015; Bay-Williams, 2016), more recent surveys show that most teachers feel prepared to teach the CCSS (Scholastic, 2014), have acquired good or excellent knowledge of the CCSS (Kane et al., 2016), base their curricula on the CCSS (Opfer et al., 2016), and use textbooks based on the CCSS (Blazar et al., 2019). Still, the challenges surrounding the implementation of the CCSS warrant robustness checks of the treatment coding in which we base the definition of students' CCSS exposure on the implementation of the CCSS, not just its adoption.

To account for this issue, we draw on a variety of data sources to create alternative treatment indicators based on CCSS implementation. They incorporate information on states' legal CCSS implementation requirements, actual CCSS implementation strategies, effectiveness of CCSS implementation, temporal CCSS implementation, and CCSS-aligned standardized testing. Detailed explanations of the different treatment indicators and their data sources are provided when reporting the corresponding results in Section 4.6 on robustness.

4.3 Identification Strategy

To estimate the effect of the CCSS on student achievement in non-targeted subjects, we run a two-way fixed effects difference-in-differences model (DD), and several extensions including models with state-specific time trends and a triple-difference model (DDD). The DD model takes advantage of the fact that some states did not adopt the CCSS. This approach

[7] Minnesota is a special case as it did not adopt the CCSS for math, but for ELA, which we code accordingly.

builds on the idea that states without reforms in a given year act as counterfactuals for states with reforms in that year, after accounting for time-invariant differences between states and national differences between years. To capture this idea econometrically, we estimate a DD model as follows:

$$T_{istuv} = \beta * CCSS_Exposure_{istuv} + \gamma * X_{istuv} + \mu_s + \lambda_t + \theta_u + \kappa_v + \epsilon_{istuv} \quad (4.1)$$

where T_{istuv} captures standardized student achievement of student i who goes to public school in state s, and takes the test in year t, grade u and subject v. Our main estimates pool all subjects that are not targeted by the CCSS, across all available grade levels. The treatment parameter $CCSS_Exposure_{istuv}$ captures the dosage of CCSS exposure of student i attending public school in state s, and taking the test in year t, grade u and subject v. Unless noted otherwise, it is defined as the share of schooling years in which a student was exposed to the CCSS (at the time of the test).[8] β is the parameter of interest capturing the effect on student achievement of being exposed to the CCSS for the entire school career until the test date (exposure=1) relative to never being exposed to the CCSS until the test date (exposure=0).[9] In our preferred treatment coding, we define a year in a given state as exposing a student to the CCSS if the state had adopted the CCSS permanently before that year or in the same year. In robustness checks, we employ other treatment definitions.

A vector of student-level control variables X_{istuv} includes indicator variables for gender, race/ethnicity, subsidized lunch status (indicator variable equals one if student receives free or reduced price lunch), English language learner status, disability status, parental education, and home possessions (separate indicator variables for computer and books). State fixed effects μ_s, test year fixed effects λ_t, grade fixed effects θ_u, subject fixed effects κ_v, and an error term ϵ_{istuv} complete the model. Note that test year and grade jointly define each cohort. Throughout the paper, all standard errors are clustered at the state level to account for potential correlation of error terms across years within states. Regressions are weighted to be population representative. We run the main DD estimations on a sample of students attending public schools (district and charter schools) only, as the implementation of the CCSS was never mandatory for private schools.

[8] It has the same domain (between 0 and 1) for students of different grades, making effect sizes of students from different grades comparable. However, this comparability comes at the price of assuming that CCSS exposure shares can be compared across grades, i.e. that 1 year of CCSS exposure for a student in grade 4 is comparable to 2 (4) years of CCSS exposure for a student in grade 8 (12). In Section 4.6, we show robustness checks for different treatment definitions.

[9] The cross-sectional data structure of NAEP only allows us to observe students at the test date, but not to follow them through their school career. Hence, we have to assume that students did not move between states during school age before the test. To the extent that pre-test moving patterns between states are uncorrelated with CCSS adoption, this merely causes attenuation bias. In that sense, our effects should be interpreted as lower-bound estimates.

This baseline model addresses a variety of concerns about our ability to estimate the causal effect of CCSS exposure on student achievement in non-targeted subjects. First, one might be worried that state-level differences in domains such as returns to education, cultural characteristics that promote educational success, genetic endowments, or preferences for centralizing policies are correlated with CCSS exposure and affect student achievement. The state fixed effects eliminate all constant differences between states. Hence, we exploit cross-cohort variation within states. Second, one might be concerned that national trends in student achievement, for example fueled by overall economic development or national education policies, appear as effects of CCSS exposure. However, our year fixed effects capture all variation in student achievement that occurs nationwide between years. In addition, our individual-level control variables ensure that the students we compare are similar with regards to demographic and socioeconomic characteristics. For these reasons, our DD model yields a causal effect of CCSS exposure on student achievement in non-targeted subjects if the main identifying assumption about parallel trends holds. It assumes that in the absence of states adopting the CCSS, the change in student achievement in treated states would have been the same as that in non-treated states.

Although this assumption cannot be directly tested, we perform a series of robustness checks to assess its plausibility. We begin with running non-parametric event-study specifications, in which the adoption of the CCSS in a given state and year is defined as the event. In contrast to the DD model, the event-study model can assess non-linear pre-reform trends in student achievement. If student achievement prior to the adoption of the CCSS was trending in the direction of the estimated CCSS effects, this could indicate a bias from underlying trends in the data. Another advantage of the event-study model is that the time course of effects of the adoption of the CCSS can be assessed by disentangling effects which occur directly at the time of the CCSS adoption from those which occur gradually after the CCSS adoption. Specifically, we estimate the effect of the CCSS adoption in year t_s on student achievement k years before and after CCSS adoption, as captured by the parameter vector β_k, see equation (4.2). These effects are estimated relative to the year of reform k=0.[10]

$$T_{istuv} = \sum_{k=-6}^{6} 1(t_{is} = t_s + k)\beta_k + \gamma * X_{istuv} + \mu_s + \lambda_t + \theta_u + \kappa_v + \epsilon_{istuv} \qquad (4.2)$$

Although the non-parametric specification captures an overall pre-trend, it does not account for state-specific trends. To address this issue, we perform analyses in which we add state-specific linear time trends, linear and quadratic time trends, as well as linear, quadratic and cubic time trends to equation (4.1). The state-specific linear time trend variable interacts each state fixed effect with a re-scaled year variable that equals one in the first year of observation,

[10] To smooth the numbers of observations across years, the observations are grouped together to bins of 2 years for all pre- and post reform years except for the bins at the beginning (end) of the domain which additionally include the years prior to (following) the domain's starting (ending) year.

two in the second year of observation, and so forth. The corresponding reform effect is identified from within-state deviations in student achievement from smooth linear trends that coincide with the different timing of CCSS adoption across states. State-specific quadratic and cubic time trends are defined analogously. Taken together, the event-study model and the models with state-specific trends address concerns about underlying trends in the data.

Even if there are no underlying trends in the data, shocks or events that occur simultaneously to the adoption of the CCSS remain a threat to the parallel trends assumption. To address this issue, we first run a triple difference model (DDD). For this analysis, we add an additional control group of private school students to our DD sample of public school students. The CCSS has never been mandatory for private schools. Correspondingly, we code private school students as not being exposed to the CCSS, even if their school is located in a state that has adopted the CCSS for public schools in some years of the school career of the students. Given the possibility that some private schools might have voluntarily implemented some elements of the CCSS, we should interpret DDD effects as lower-bound estimates,[11] (at least under the assumption of no endogenous selection between public and private school students as we discuss below).

Econometrically, we capture our third difference using a school type indicator variable (public school vs. private school). The full DDD follows equation (4.1) but adds a baseline indicator for school type as well as a full set of fixed effects interactions. This set includes school-type-by-state fixed effects (for example to control for state-specific time-constant regulation differences between private and public schools), school-type-by-year fixed effects (for example to control for changes in the national funding of public schools), and state-by-year fixed effects (for example to control for state-specific policies and programs directed at students or their families regardless of school type). The DDD uses variation at the school-type-by-state-by-year level to identify the effect of the CCSS on student achievement from differences in student achievement of students who attend public school compared to student achievement of students who attend private school coincident with the timing of the CCSS adoption in each state. The identifying assumption of the DDD requires that there is no other school type specific variable correlated with the CCSS adoption that affects student achievement. This identifying assumption is substantially weaker than that of the DD model, as it cancels out all confounding variables that affect public and private school students equally.

Still, policies that occurred simultaneously to the adoption of the CCSS which affect public and private school students differently and which have an effect on student achievement could bias the models presented so far. To address this concern, we collect data on reforms of public schooling policies and private schooling policies. Examples of public schooling policy controls include waivers from NCLB/ESSA accountability requirements, the adoption of the

[11] If some private schools implemented elements of the CCSS, we would code some students as untreated although they have in fact received some treatment. Hence, we would erroneously difference out some part of the real effect of the CCSS.

Next Generation Science Standards, and public education expenditures (as measured by the district-by-year-level per-pupil education expenditures in logarithmized dollars). Examples of private schooling policy controls include states' control of private school licensure, of private school curricula, or publicly funded voucher laws. Similarly, we also add controls for policies on homeschooling and compulsory schooling as a robustness check, as they might indirectly and differently affect public and private schools. The extent of recordkeeping requirements for homeschooling is an example of a homeschooling policy, and the number of compulsory schooling years is an example of a compulsory schooling policy. Adding these policy variables as controls can alleviate many concerns about simultaneous policies biasing the results.

Another threat to validity is that the adoption of the CCSS might have caused heterogeneous selection of specific groups of students into school types. For example, estimates could be biased if students with politically conservative parents left the public school system at the same time that the CCSS was adopted, and if political conservatism of parents affects student achievement. Neither the DDD model, nor the DD models based on a sample of public school students, are immune to this selection issue. We address this concern by running a DD model with a joint sample of public and private school students, whose reform effects are net of any heterogeneous selection between school types.

Even if the parallel trends assumption holds, the previously presented two-way fixed effects models could yield biased results due to time-varying treatment effects. If already-treated students act as controls for later-treated students in settings with staggered treatment timing, time-varying treatment effects can bias results away from the true effect. This issue, also referred to as "negative weighting", has received much attention in the recent econometric literature (Chaisemartin and D'Haultfœuille, 2020; Athey and Imbens, 2021; Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Roth and Sant'Anna, 2021; Sun and Abraham, 2021). We address this issue by conducting a robustness check in which we exclude all 2x2 DD comparisons from the sample in which already-treated students act as controls. In sum, each of the presented approaches in this section has different identifying assumptions and addresses concerns about underlying trends in the data, simultaneous shocks, selection, or negative weights, in a different way. In our view, robust insights about the effect of the CCSS on student achievement in non-targeted subjects can be obtained if the different approaches yield similar results.

4.4 Data

4.4.1 Student Achievement Data

We merge data on the adoption and implementation of the CCSS with standardized student achievement data. The data sources on the adoption and implementation of the CCSS are described in Section 4.2 on the institutional background and in Appendix A4.3. For student achievement, we use the restricted-use individual-level dataset of the NAEP. The NAEP is a

congressionally mandated project which is nationally representative of the US student body. It is also known as The Nation's Report Card. NAEP has measured the knowledge of US students in various subjects since its first assessment was conducted in 1969.[12] The assessments are administered by the National Center for Education Statistics (NCES), an institution within the Institute of Education Sciences (IES) and the US Department of Education. Notably, there is a significant overlap between CCSS and NAEP items (Daro et al., 2015). For example, 79 percent of items on grade 4 and 87 percent of items on grade 8 of the 2015 NAEP math assessment were also covered by the CCSS.

The NAEP individual-level dataset has several advantages for our analysis. First, it provides information on student achievement at the individual level for the relevant years, grades, and subjects for all US states. Its comparability across states and over time allows for consistent standardization and two-way fixed effects difference-in-differences estimations. Specifically, we use the NAEP data from 2005 to 2015 for all available grade levels, namely grades 4, 8 and 12.[13] We exclude data from 2004 and before, as NAEP sampling increased tremendously after 2001 (and no testing was done in non-targeted subjects between 2002 and 2004 and after 2015). Besides, this sample cut provides pre- and post reform periods of roughly equal duration. We use the NAEP data from science, civics, economics, geography, and history to capture student achievement in non-targeted subjects.[14] Table A4.1 lists the grades in which the NAEP tests were administered for each of the five subjects for each year between 2005 and 2015 (and for which state identifiers of students are available). We use the student-level data from all these subject-year-grade combinations in our main analysis. The resulting sample consists of more than one million students.[15] [16]

Second, the NAEP data include a rich set of individual-level control variables such as student gender, race/ethnicity, and various socio-economic background variables, among others. Hence, we can control flexibly for students' pre-reform characteristics and perform subgroup analyses. We set missing values of controls to zero and add separate explanatory binary variables to all regressions to account for these missing values, unless noted otherwise.[17] Third, the NAEP also administers student achievement tests and surveys in private schools, which we can exploit for identification given that the CCSS was never mandatory for private

[12] Throughout the paper we use the Main-NAEP and not the Long Term Trend NAEP, as the Main-NAEP has much larger sample sizes and is state-representative.

[13] For the targeted subjects evaluated in Appendix A4.2, we have data until 2017.

[14] Other tests such as theater, visual arts, and music were not tested in the relevant years.

[15] Throughout the paper, we report the number of observations rounded to the nearest ten digit to comply with data protection regulations of the NCES.

[16] In the NAEP, no student takes the entire student achievement test. Instead, the NAEP reports plausible values for overall student achievement on a test estimated from the sample of questions that were administered to a student. We make the arbitrary choice of selecting the second plausible value the NAEP provides. The raw correlation between the second plausible value and the average of the first five plausible for student achievement in, for example, science equals 0.95 in our sample. Our results are robust to estimating effects using any other plausible value.

[17] Our results are robust to not imputing the missings, see Table A4.10.

schools. Fourth, the NAEP has also surveyed the teachers of a subset of the students in the sample. We can make use of the teacher data to investigate changes in classroom instruction resulting from the CCSS, as presented in Section 4.7 on mechanisms.

4.4.2 Descriptive Statistics

Table A4.2 presents the mean, standard deviation, minimum and maximum of the main variables. The main outcome variable is student achievement, which we standardize to have a mean of zero and a standard deviation of one in the first year of each available grade-subject combination, following (Lafortune et al., 2018; Jackson et al., 2021). This standardization allows the mean and variance to flexibly evolve in the following years, which explains why their overall mean and standard deviation reported in Table A4.2 are close but not equal to zero and one, respectively. Regarding student characteristics, we note that about half of the sample is female and almost 60 percent is White. The shares of Black and Hispanic students are 15 percent and 20 percent, respectively. 5 percent of the students in the sample are Asian students (including Pacific Islanders). About 6 percent of the sample have English language learner status, and 11 percent disability status. To assess what share of students come from a low socioeconomic background, we can look at the shares of students receiving subsidized lunch (43 percent), having parents who did not finish high school (8 percent), having no computer at home (10 percent), or having less than 10 books at home (13 percent). We also show descriptive statistics for variables measuring the instructional focus on non-targeted subjects and teacher characteristics (both based on the NAEP teacher surveys), which we analyze in more detail in Section 4.7.

4.5 Results

4.5.1 Main Results

Table 4.1 presents estimates of the statistical relation between CCSS exposure and student achievement in non-targeted subjects for different sets of control variables. In column 1, there are no control variables. The positive and statistically significant correlation between CCSS exposure and student achievement in non-targeted subjects implies that students exposed to the CCSS perform better in non-targeted subjects than those not exposed to the CCSS. This correlation could be caused by CCSS exposure improving student achievement in non-targeted subjects, for example through positive spillovers. It could also be caused by above-average student achievement in non-targeted subjects leading to CCSS exposure. This reverse causality could occur, for example, if above-average student achievement in non-targeted subjects before the reform promotes confidence in national education policies, thereby encouraging states to adopt the CCSS. Moreover, the observed positive correlation between CCSS exposure and student achievement in non-targeted subjects could also be driven by third variables such as parental education. This would be the case, for example, if states with a high proportion of

students with highly educated parents are more likely to adopt the CCSS, for instance if these parents vote disproportionately for parties that endorse the CCSS, and if parental education itself increases student achievement in non-targeted subjects.

To isolate the effect of CCSS exposure on student achievement, we add control variables in columns 2-3. The positive correlation of CCSS exposure and student achievement in CCSS subjects remains almost unchanged conditional on student-level control variables (column 2). In the full model with both student-level controls as well as state and year fixed effects, the positive correlation becomes negative (column 3). The full model is our preferred model, as it flexibly accounts for demographic and socioeconomic differences between students as well as time-invariant differences between states and national differences between years. Unless noted otherwise, all further models presented in this paper are full models.

More specifically, exposure to the CCSS during the entire school career (at the time of testing), as opposed to no CCSS exposure at all, decreases student achievement in non-targeted subjects on average by 0.08 units of a standard deviation. This effect is statistically significant at the 5 percent level. To illustrate the effect size, we draw on the literature on education production functions, which suggests that the gain in learning from one year of schooling is equivalent to about one-quarter to one-third of a standard deviation increase in student performance on standardized tests (Woessmann, 2016). Correspondingly, the CCSS-induced learning loss in non-targeted subjects is equivalent to approximately 25 percent to 30 percent of a school year.

4.5.2 Subgroup Analysis

The negative effect of the CCSS on student achievement in non-targeted subjects need not be evenly distributed across subgroups. We hypothesize that students from underprivileged backgrounds may be disproportionately disadvantaged as it is more difficult for themselves, their parents, or their social environments in general to compensate for the reduction of instructional focus on non-targeted subjects.[18] For example, parents from underprivileged backgrounds might be less able or might have less time to help their children with homework themselves or pay for private tuition. To test this hypothesis, we conduct subgroup analyses by students' demographic and socioeconomic characteristics.

As reported in Table 4.2, we find that the negative effect on student achievement in non-targeted subjects is not evenly distributed across subgroups. With respect to race/ethnicity, the student achievement of Hispanics and Blacks in non-targeted subjects is reduced disproportionately as a consequence of the CCSS, while there are almost no reform effects for Whites and Asians. Race and ethnicity aside, the negative effect is larger for students who qualify for subsidized lunch than for those who do not. Furthermore, students with English

[18] In Section 4.7 on mechanisms, we use outcomes measuring instructional focus based on teacher survey data to show explicitly that the CCSS caused a reduction of instructional focus on non-targeted subjects.

language learner status or disability status lose disproportionately. Taken together, these results indicate that students from groups typically regarded as socially, economically or physically underprivileged suffer most from CCSS exposure in terms of their achievement in non-targeted subjects.

We also perform subgroup analyses by subjects and grades. As far as subjects are concerned, the negative effect on student achievement in non-targeted subjects comes mostly from science, as we show in Table A4.3. In terms of grades, the negative effect is mostly due to students from grade 4, see Table A4.4. These subgroup effects should be interpreted with caution as testing frequencies and sample sizes are much larger for science (compared to civics, economics, geography, and history) for grades 4 and 8 (compared to grade 12). Still, the large subgroup effect for students in grade 4 makes intuitive sense as teachers in elementary school have the greatest flexibility in shifting the instructional focus.[19]

4.6 Robustness

In this section, we test the robustness of the main result to account for four general types of concerns. First, we test the plausibility of the identifying assumption about parallel trends between treatment and control groups in the two-way fixed effects DD model in various ways in a series of econometric robustness tests. Second, we assess whether negative weighting induced by time-varying treatment effects affects our two-way fixed effects DD estimates and their interpretation. Third, we test whether re-defining our treatment variable to incorporate information about CCSS implementation changes our results. Fourth, we conduct a series of further specification checks to ensure that our results do not hinge on arguably arbitrary specification choices in the main regression.

4.6.1 Robustness Tests on Parallel Trends Assumption

A first plausibility test for the identifying assumption of parallel trends in outcomes between treatment and control groups are event-study specifications. Here, the adoption of the CCSS in a given state and year is defined as an event. As depicted in Figure 4.2, no statistically significant pre-trend in student achievement in non-targeted subjects can be identified prior to the adoption of the CCSS. If at all, student achievement in these subjects was improving before the reform. In contrast to this insignificant positive pre-trend, student achievement

[19] Yet another subgroup analysis we perform is motivated by the hypothesis that states with high average student achievement levels before the adoption of CCSS might suffer more from CCSS adoption in terms of student achievement in non-targeted subjects relative to states that had been low-performing to begin with. To test this hypothesis, we perform subgroup analysis by quartiles of the states' pre-CCSS student achievement level. As presented in Table A4.5, there is no linear effect pattern across quartiles. Although the effect difference between the lowest quartile and the highest quartile is negative which could support our hypothesis, we do not emphasize this subgroup finding as it does not hold in robustness checks on the econometric and treatment specifications as described in Section 4.6.

declined substantially after the reform, both in terms of size and significance. This finding supports the validity of the parallel trends assumption and confirms the negative effect of the CCSS on student achievement in non-targeted subjects reported in the main analysis.

To further account for the possibility of state-specific trends and simultaneous shocks, Table 4.3 presents econometric robustness tests for the main regression results (shown again in column 1 to facilitate comparison) by first adding state-specific linear time trends (column 2), then adding state-specific quadratic time trends (column 3), and finally adding state-specific cubic time trends (column 4). In column 5, DDD estimates are reported in which a set of data on private school students is added to the sample. Here, all state-specific time trends are replaced with a fixed effect for school type (public school vs. private school), as well as with school-type-by-state, school-type-by-year, and state-by-year fixed effects. In column 6, we run the main two-way fixed effects DD model with a sample of all students from public and private schools.

We find that the negative and significant main effect is robust across specifications and becomes even slightly more negative in the models with state-specific time trends of various orders. The same is true for the DDD model. The model which is estimated based on the sample of both public and private school students yields an effect that is slightly smaller than that of the main model, but still negative and significant. This series of econometric robustness checks demonstrates that the main effects are not affected by underlying state-specific trends in the data, endogenous selection between public and private school students, and shocks that occur simultaneously to the adoption of the CCSS as long as they affect both public and private school students equally.

While a large variety of state-specific shocks, including policy reforms in many areas, plausibly affect public and private schools students equally, the parallel trends assumption could still be violated by state-specific policies for public and private schooling. To test for this possibility, we perform robustness checks which explicitly control for state-specific time-varying policies for public schooling (Table A4.6) and private schooling (Table A4.7), respectively. The results are robust throughout. The first, second, and third specification in Table A4.6 controls for NCLB waivers, state proficiency standards, and the adoption of the Next Generation Science Standards, respectively. The robustness of the CCSS exposure coefficient supports aforementioned interpretation that it is indeed the CCSS itself that drives the reported effects, and not waivers, other proficiency reforms or the NGSS. In addition, policies on homeschooling and compulsory schooling could indirectly affect public and private schooling decisions. To account for this possibility, we also present robustness checks in which we add controls for homeschooling and compulsory schooling. As can be seen in Table A4.8 for homeschooling and Table A4.9 for compulsory schooling, the results are robust. Taken together, the findings in this subsection indicate that neither underlying trends in the data nor shocks occurring simultaneously to the adoption of the CCSS give rise to our main result.

4.6.2 Robustness Tests on Time-Varying Treatment Effects

Another potential threat to identification of the presented two-way fixed effects DD models are time-varying treatment effects. The main estimate is a weighted sum of the average treatment effects in each state and year (i.e. of each 2x2 comparison), with weights that may be negative. These negative weights can cause the main regression coefficient to be negative although all the average treatment effects are positive. Weights can be negative if already-treated units act as controls for later-treated units, in settings with time-varying treatment effects and staggered reform adoption.

Our event-study graphs presented above provide first insight into the potential bias induced by time-varying treatment effects, as they allow us to separate instantaneous from gradual reform effects. Event-studies are immune to bias from time-varying treatment effects as long as the pattern of effects is the same for all treatment cohorts. To explicitly explore the issue of time-varying treatment effects, we create a sample in which already-treated students never act as controls. Creating this sample is relatively straightforward in our setting as most states adopted the CCSS in 2010 or did not adopt the CCSS at all. By excluding the six states which adopted the CCSS in 2011 and 2012 from the sample, we transform our staggered setting into a non-staggered setting that is immune to negative weights. As shown in column 1 of Table A4.10, the negative significant effect of the CCSS on student achievement in non-targeted subjects remains in this modified sample. This finding demonstrates that the main result is not driven by time-varying treatment effects and negative weights.

4.6.3 Robustness Tests on Treatment Definition

A different type of concern is that CCSS adoption and CCSS implementation could diverge. In our preferred treatment coding, we count all years as causing CCSS exposure for a student in a given state, in which the state had permanently adopted the CCSS before that year or at most in the same year. However, states that have adopted the CCSS permanently may not have implemented the CCSS comprehensively and thus may not be creating actual exposure. Conversely, states that have not adopted the CCSS permanently may have adopted and/or implemented the CCSS temporarily or partially.

To test whether our results hold if we define treatment based on CCSS implementation, we re-run our main regression using five different treatment variables, each capturing different information about the implementation of the CCSS. Under these treatment definitions, a school year in a given state is defined as a school year with CCSS exposure if that state (i) expects teachers to fully incorporate the CCSS in their classroom instruction, (ii) followed at least two out of three CCSS implementation strategies (professional development, new instructional materials, joined testing consortium), (iii) observed an effective change in state standard content due to the adoption of the CCSS, which we define to mean that no state standard existed that closely resembles the CCSS before the adoption of the CCSS, (iv) adopted and/or implemented the CCSS at least temporarily, or (v) mandated standardized tests aligned

to the CCSS. Further information on each treatment definition, its construction, data sources, including a table containing state-specific coding information for all treatment definitions are provided in Appendix A4.3.

We present the results for each treatment definition (including our main result to facilitate comparison) in Table 4.4. We find negative point estimates throughout, ranging from -0.088 to -0.035 units of a standard deviation for the overall sample. However, most of them are statistically insignificant. To assess this finding further, we also show effects for the subsample of students in grade 4 for which we have observed the largest subgroup effects in the main analysis. At least for this hand-picked subsample, we find statistically significant effects throughout (and point estimates ranging from -0.177 to -0.098 units of a standard deviation). Taken together, these findings suggest that results using treatment definitions based on CCSS implementation rather than CCSS adoption lead to the same overall conclusion as the main results in Section 4.5.

4.6.4 Further Specification Checks

In addition, we want to assess whether our results are robust to a number of modifications of our main regression. As indicated before, we set missing values of controls to zero and add separate explanatory binary variables to account for these missing values in our main regressions. The shares of missing values for the student control variables are below 10 percent for all variables except for parental education. For the latter approximately 40 percent of the values are missing, which can be mostly explained by the fact that this question was not asked in grade 4. To test whether the parental education control and its imputation affect the results, we run our main regression without controlling for parental education. As shown in column 2 of Table A4.10, the effects do not differ meaningfully. As an additional robustness check, we do not impute missing values of any control variables (in addition to leaving parental education out of the set of control variables). As can be seen in column 3, the results are robust.

Moreover, we test the robustness of our main regression by modifying the definition of the treatment variable that captures the dosage of a student's exposure to the CCSS. So far, we have defined this variable as the share of schooling years a student was exposed to the CCSS (at the time of the survey). Alternatively, we now define exposure to the CCSS as the number of schooling years a student was exposed to the CCSS (at the time of the survey). As shown in column 4 of Table A4.10, the negative effect is now insignificant and much smaller, but has a similar interpretation. In particular, we find that a one-year increase in CCSS exposure reduces student achievement in non-targeted subjects by 0.006 units of a standard deviation. Assuming 12 years of schooling, the total effect of CCSS exposure throughout the entire school career, as opposed to no exposure, equals 0.072 units of a standard deviation (0.006*12). This number is close to the result of our main regression (0.079 units of a standard deviation) in which we define the treatment variable as a share of years. In addition, we show that our

results are robust to excluding charter schools from the sample of public schools, or omitting population weights, respectively, see columns 5 and 6.

4.7 Mechanisms

To study what gave rise to the observed effect on student achievement in non-targeted subjects, we examine what changed in students' classrooms in these subjects due to the CCSS. To this end, we draw on teacher survey data, provided by the NAEP for a subset of waves and classrooms. This data is suitable for our analysis for several reasons. First, it contains a rich set of subject-specific questions on instructional focus in the classroom comprising instruction time, instructional resources, five measures of differentiated instruction, and four measures of the quality of teacher-student interactions. We note that the instructional focus outcomes could be endogenous to the reform and hence should be interpreted as changes in teachers' perceptions of classroom instruction rather than evidence based on administrative data (which is not available for these outcomes at the subject-state-year-level).

Second, the NAEP includes teacher background characteristics which we can use as control variables and for subgroup analyses. Third, the NAEP teacher surveys are linked to the NAEP student achievement tests and student surveys. This link allows us to examine how instructional practices changed in non-targeted subjects according to the teachers who taught precisely the tested students from our main analysis. Fourth, the teacher surveys are standardized in the same way as the student surveys and achievement tests, making them comparable across states and years and thus suitable for a two-way fixed effects difference-in-differences approach. In fact, we can keep the empirical framework from the previous sections largely unchanged, but we use instructional focus outcomes instead of student outcomes and add teacher controls, thus ensuring methodological consistency with the previous sections.

Table 4.5 presents the results of CCSS exposure for instructional focus in non-targeted subjects. Overall, we find that the CCSS caused a reduction in instructional focus on the non-targeted subjects. Specifically, we observe negative significant effects of the CCSS on weekly instruction time, provision of instructional materials and resources, and two dimensions of the quality of teacher-student interactions. These two dimensions are setting and discussing goals with students. The extent of differentiated instruction did not change meaningfully. To illustrate the interpretation of the reported point estimates, we note that teachers of students who are fully exposed to the CCSS are 17 percentage points less likely to teach these students more than five hours per week in non-targeted subjects than teachers of students with no CCSS exposure, conditional on teacher characteristics, student characteristics as well as state, year, grade and subject fixed effects.[20]

[20] The answer categories of the instructional outcome variables were coded differently in different survey waves of the NAEP. Hence, we code the variables as reported in the footnote of Table 4.5 to ensure consistency across waves.

The reduction in instructional focus on non-targeted subjects does not have to be evenly distributed across teacher subgroups. Understanding which subgroups of teachers drive the effects is interesting in itself and can be useful for tailoring policy advice to specific groups of teachers. We perform subgroup analyses for the four instructional focus outcome variables which were most affected by the CCSS, namely instruction time, instructional resources, and teacher-student interactions (setting and discussing goals). We conduct subgroup analyses by teacher characteristics which include teacher race/ethnicity, teacher education, teacher certification, and teacher experience. The subgroup pattern is not evenly distributed across instructional focus outcomes, but in general we find the largest reductions in instructional focus on non-targeted subjects for White teachers and teachers without a certification, see Tables A4.11, A4.12, A4.13 and A4.14, respectively for the four instructional focus outcomes.

Altogether, these results show that the adoption of the CCSS has shifted the instructional focus away from the non-targeted subjects. This finding is in line with the results from Section 4.5, which show a decline in student achievement in these subjects. It is also consistent with previous literature showing that instructional inputs affect student achievement. Increases in instruction time (Taylor, 2014), instructional resources (Holden, 2016), and the quality of teacher-student interactions (Allen et al., 2011) have all been shown to positively affect student achievement. These instructional inputs can also interact. For example, the effect of instruction time on student achievement depends on student-teacher interactions (Rivkin and Schiman, 2015).

4.8 Conclusion

Since 2010, the majority of US states have aligned their education standards for math and ELA by adopting the CCSS. This paper estimates the effect of CCSS adoption on student achievement in non-targeted subjects. We find that the CCSS reduced student achievement in non-targeted subjects, particularly for younger underprivileged students in science. Student achievement is a relevant outcome in its own right, but also affects long-term societal efficiency (Hanushek and Woessmann, 2008, 2012) and equity (Fryer, 2011). In sum, our results allow to evaluate the CCSS more comprehensively, with, at best, modest positive effects on student achievement in targeted subjects (also previously documented in Bleiberg (2021)) at the expense of student achievement in non-targeted subjects.

With respect to mechanisms, we find that the negative spillover of the CCSS on student achievement in non-targeted subjects was accompanied by a reduction of instructional focus on these subjects. This result mirrors previous findings on the effects of NCLB, which also only focused on math and ELA and caused a reduction in instruction time in non-targeted subjects (Reback et al., 2014).[21] However, in contrast to the effects of the CCSS documented in our

[21] Anecdotal evidence attributes this to NCLB's measure of adequate yearly progress (AYP) for schools which has been determined through annual growth in reading and mathematics. As a result, instruction time and

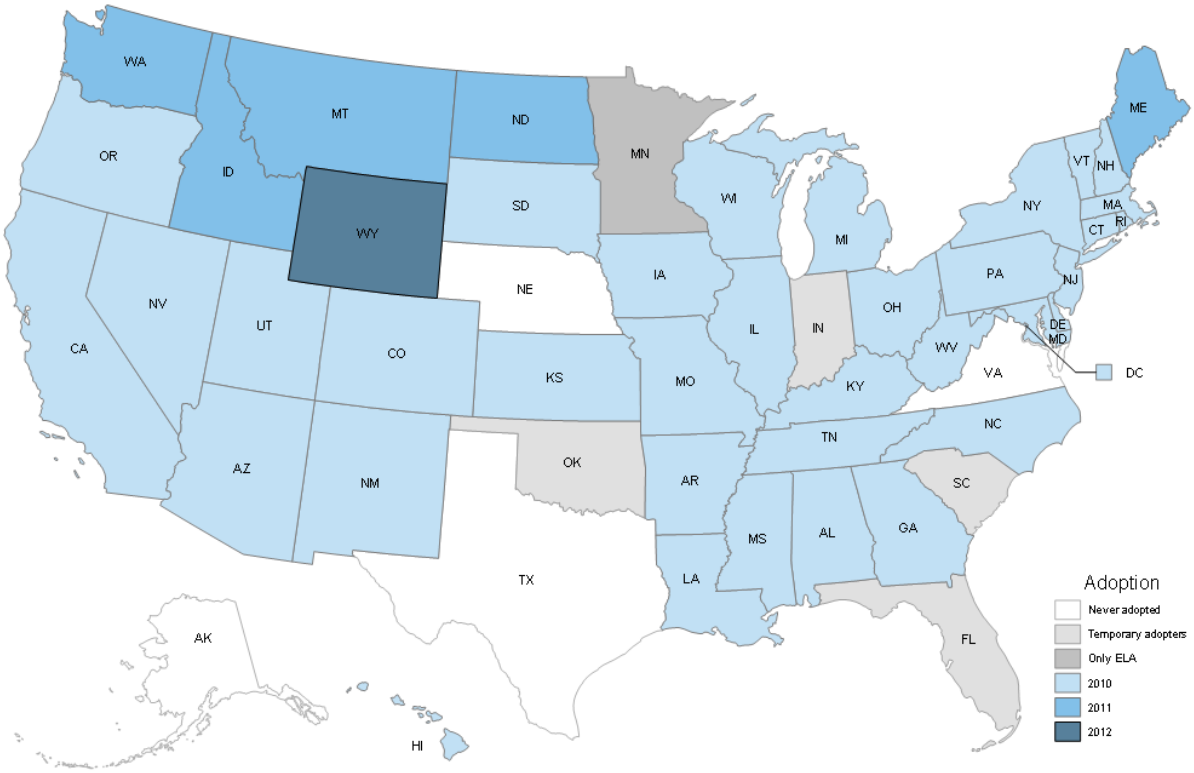
paper, NCLB did not did not decrease student achievement in non-targeted subjects (Dee and Jacob, 2011; Reback et al., 2014). Potentially, these differential effects on student achievement in non-targeted subjects are rooted in the accountability pressures of NCLB that led to much more pronounced student achievement gains in the targeted subjects relative to CCSS. These gains may have created positive spillovers on student achievement in non-targeted subjects, and may have made shifts in instructional focus matter less. In any case, the results presented in this paper extend the literature on strategic reactions of educators to reforms to the topic of education standards.

In terms of education policy, our results might imply that the negative spillover of the CCSS on student achievement in the non-targeted subjects could have been avoided if more subjects had been included in the CCSS. Arguably, such a comprehensive education standard could have prevented that some subjects are deemed less relevant and receive less instructional attention. At the same time, such a policy might also have reduced any positive effects on student achievement in the targeted subjects. However, adopting a centralized education standard which covers all subjects requires that the participating states agree on the educational content for each subject. To achieve this goal, political challenges need to be overcome as exemplified by the controversies around the history curriculum (Cohen, 2020) or around the treatment of evolution theory in US State Science Education Standards (Lerner, 2000a; Arold, 2022).

other dimensions of instructional focus such as teacher practice support have been reduced for science and social studies (see Settlage and Meadows (2002) and Johnson (2007) for science, and Misco (2005) and National Council for Social Studies (2008) for social studies).

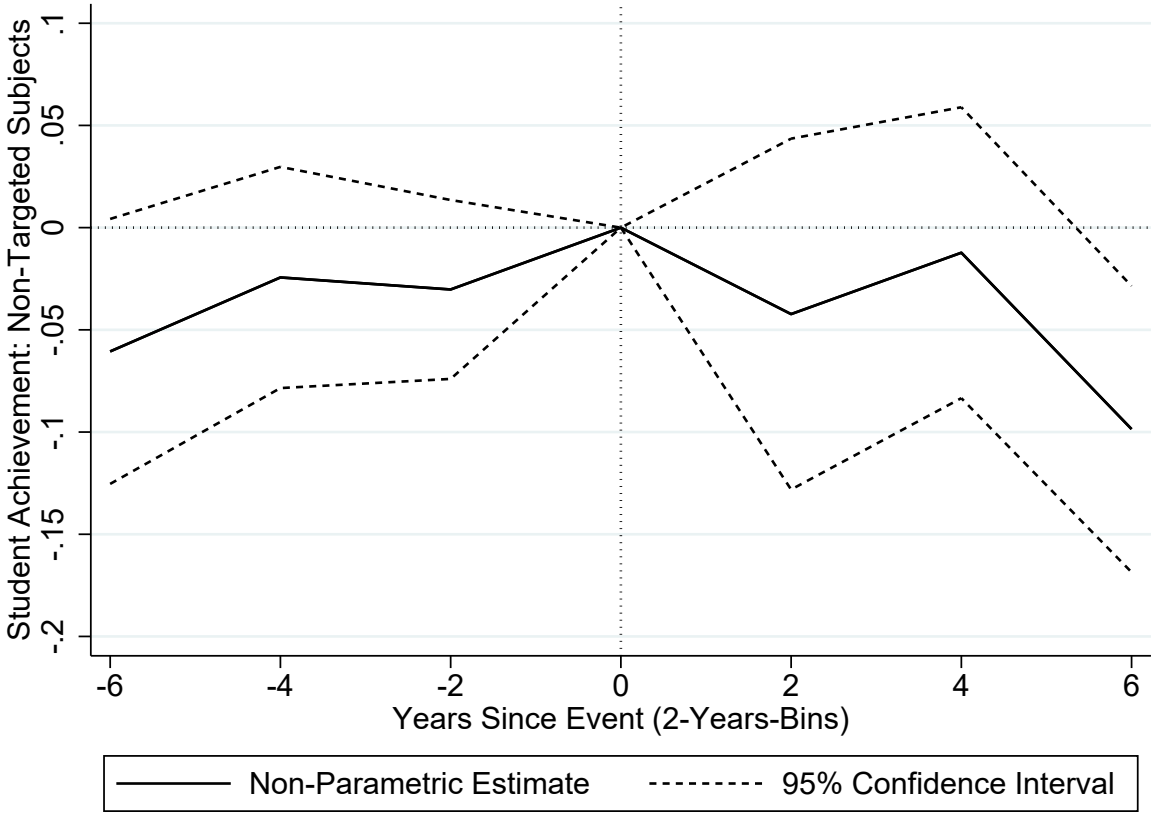
Figures and Tables

Figure 4.1: CCSS adoption map



Notes: Map depicts state-level adoption of CCSS. Data sources: Achieve Inc. (2013), Bleiberg (2021), and CCSSI (2022)

Figure 4.2: Event-study graph: Non-targeted subjects



Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as state, test year, grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Numbers on horizontal axis refer to respective two-year bins; i.e. 2 = first two years of treatment (year 0 = excluded category). The p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.312 and 0.001, respectively. Data source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress

Table 4.1: Effect of CCSS exposure on student achievement in non-targeted subjects

	(1)	(2)	(3)
CCSS Exposure	0.105***	0.117***	-0.079**
	(0.036)	(0.043)	(0.036)
State and Year FEs	NO	NO	YES
Controls	NO	YES	YES
Adj. R-squared	0.001	0.379	0.390
Observations	1,103,630	1,103,630	1,103,630

Note: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table 4.2: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by student characteristics

	Gender		Race/Ethnicity					Subsidized Lunch Status		English Languange Lerner Status		Disability Status	
	(1) Female	(2) Male	(3) White	(4) Black	(5) Hispanic	(6) Asian	(7) Other	(8) Yes	(9) No	(10) Yes	(11) No	(12) Yes	(13) No
CCSS Exposure	-0.079**	-0.080**	-0.016	-0.111***	-0.181***	-0.012	0.016	-0.096***	-0.041	-0.201***	-0.065*	-0.176***	-0.068*
	(0.037)	(0.036)	(0.039)	(0.037)	(0.041)	(0.061)	(0.071)	(0.031)	(0.044)	(0.073)	(0.035)	(0.049)	(0.037)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.395	0.390	0.272	0.241	0.331	0.392	0.362	0.303	0.301	0.182	0.348	0.308	0.353
Observations	544,410	559,210	631,640	184,680	194,250	55,130	37,930	513,910	589,710	69,190	1,034,440	128,590	975,030

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table 4.3: Effect of CCSS exposure on student achievement in non-targeted subjects, econometric robustness

	(1) DD	(2) (1) + linear state trends	(3) (2) + quadratic state trends	(4) (3) + cubic state trends	(5) DDD	(6) DD with private schools
CCSS Exposure	-0.079**	-0.117**	-0.100**	-0.095**	-0.090**	-0.074**
	(0.036)	(0.044)	(0.047)	(0.045)	(0.044)	(0.033)
State and Year FEs	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.390	0.391	0.392	0.393	0.394	0.390
Observations	1,103,630	1,103,630	1,103,630	1,103,630	1,135,960	1,135,960

Notes: Each entry is from a separate two-way fixed effects regression model, where Model (1) is the baseline model, Models (2), (3), (4) subsequently add linear, quadratic, and cubic state-specific time trends, and Model (5) presents a triple-difference model where school type (public vs. private school students), school type*state, school type*year, and state*year fixed effects replace all state-specific time trends. Model (6) estimates the basic two-way fixed effects model on a sample of public and private school students. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table 4.4: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness using different definitions of treatment implementation

	CCSS adoption		CCSS implementation requirement		CCSS implementation strategies		Effective CCSS implementation		Include temporary CCSS adopters and implementers		CCSS-aligned testing	
	(1) All grades	(2) Only grade 4	(3) All grades	(4) Only grade 4	(5) All grades	(6) Only grade 4	(7) All grades	(8) Only grade 4	(9) All grades	(10) Only grade 4	(11) All grades	(12) Only grade 4
CCSS Exposure	-0.079**	-0.134***	-0.088	-0.177**	-0.042	-0.098*	-0.075**	-0.107***	-0.035	-0.111**	-0.085	-0.118**
	(0.036)	(0.036)	(0.079)	(0.076)	(0.041)	(0.050)	(0.035)	(0.039)	(0.045)	(0.055)	(0.057)	(0.054)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.390	0.376	0.390	0.376	0.390	0.376	0.390	0.376	0.390	0.376	0.390	0.376
Observations	1,103,630	434,440	1,103,630	434,440	1,103,630	434,440	1,103,630	434,440	1,103,630	434,440	1,103,630	434,440

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variables: Share of schooling years a student was exposed to CCSS (at the time of testing), where in Models 1 and 2 (CCSS adoption, baseline model) each schooling year counts as exposed in a given state in which the state adopted the CCSS permanently before that year or in the same year according to Achieve Inc. (2013) and CCSSI (2022), where Models 3 and 4 (CCSS implementation requirement) each schooling year counts as exposed in a given state in which the state expects teachers to fully incorporate CCSS into classroom instruction in grades K-12 in English language arts and mathematics according to Achieve Inc. (2013) and CCSSI (2022), where in Models 5 and 6 (CCSS implementation strategies) each schooling year counts as exposed in a given state if state education agency officials report that their state pursued at least two out of three CCSS implementation strategies (professional development, new instructional materials, joined testing consortium) as reported in Webber et al. (2014), where in Models 7 and 8 (Effective CCSS implementation) each schooling year counts as exposed in a given state in which the state implemented an effective change in state standard content through the adoption of CCSS which we define as not having had a state standard in place before the adoption of CCSS whose academic rigor is "too close to call" in comparison with CCSS (Carmichael et al., 2010) for the set of states adopting CCSS according to Achieve Inc. (2013) and CCSSI (2022), where in Models 9 and 10 (Include temporary CCSS adopters) each schooling year counts as exposed in a given state in which a state adopted and/or implemented CCSS at least temporarily according to Bleiberg (2021); and where in Models 11 and 12 (CCSS-aligned testing) each schooling year counts as exposed in a given state in which the state adopted CCSS-aligned standardized testing including field and transitional tests according to our own research (see Table A4.18 for state-specific details of CCSS-aligned testing). Table A4.19 provides state-specific coding information on all treatment definitions. Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table 4.5: Effect of CCSS exposure on instructional focus in non-targeted subjects

	Instruction Time	Instructional Resources	Differentiated Instruction					Teacher-Student Interactions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Standards	Material	Activities	Methods	Pace	Discuss Students' Performance	Set goals	Discuss goals	Adjust teaching
CCSS Exposure	-0.171***	-0.116***	-0.027*	0.021	-0.022	-0.015	0.006	-0.017	-0.035*	-0.047**	-0.020
	(0.055)	(0.030)	(0.015)	(0.016)	(0.013)	(0.022)	(0.016)	(0.035)	(0.019)	(0.023)	(0.023)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.090	0.043	0.053	0.073	0.064	0.167	0.122	0.109	0.091	0.089	0.058
Observations	847,830	785,660	555,490	556,350	556,070	556,440	556,110	554,820	554,490	554,360	554,250

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variables, by columns: Probability that students' teachers report about subject in question (Model 1) that students receive more than five hours of weekly instruction, (Model 2) that their school system provides them with all or most materials and other resources they need for the instruction, (Model 3) that they set differentiated standards for some students at least to a moderate extent, (Model 4) that they use differentiated materials for some students at least to a moderate extent, (Model 5) engage some students in differentiated activities at least to a moderate extent, (Model 6) that they use differentiated methods for some students at least to a moderate extent, (Model 7) change pace for some students at least to a moderate extent, (Model 8) that they discuss the student's current level of performance at least once a month, (Model 9) that they set goals for specific progress the student would like to make at least once a month, (Model 10) that they discuss progress the student has made toward goals previously set at least once a month, (Model 11) that they determine how to adjust their teaching strategies to meet the student's current learning needs and to reflect the student's future goals at least once a month. Explanatory variable: Share of schooling years teacher's students were exposed to CCSS (at the time of testing). Controls: Indicator variables for teacher race/ethnicity, teacher education, teacher certification (separate indicator variables for certification of National Board for Professional Teaching Standards, and for alternative certification), and teacher experience, as well as student controls for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Appendix

A4.1 Supplementary Figures and Tables

Table A4.1: List of grades of NAEP tests for non-targeted subjects

	Non-targeted Subjects				
Year	Science	Civics	Economics	Geography	History
2005	4, 8				
2006		4, 8, 12	12		4, 8, 12
2007					
2008					
2009	4, 8, 12				
2010		4, 8, 12		4, 8, 12	4, 8, 12
2011	8				
2012			12		
2013					
2014		8		8	8
2015	4, 8, 12				

Notes: NAEP student achievement data in non-targeted subjects at the subject-by-year-by-grade level. Data source: See Figure 4.2

Table A4.2: Descriptive statistics

	Mean	Std. Dev.	Min.	Max.
Student Achievement Outcomes:				
Science	0.08	1.03	-4.76	4.51
Civics	0.03	0.98	-4.49	3.16
Economics	0.02	0.97	-4.33	3.70
Geography	0.00	1.00	-5.15	3.90
History	0.04	0.98	-4.64	3.32
Student Controls:				
Female	0.49	0.50	0.00	1.00
Race/Ethnicity: White	0.57	0.49	0.00	1.00
Race/Ethnicity: Black	0.15	0.36	0.00	1.00
Race/Ethnicity: Hispanic	0.20	0.40	0.00	1.00
Race/Ethnicity: Asian	0.05	0.22	0.00	1.00
Race/Ethnicity: Other	0.03	0.16	0.00	1.00
English Language Learner	0.06	0.24	0.00	1.00
Disabled	0.11	0.31	0.00	1.00
Subsidized Lunch	0.43	0.50	0.00	1.00
Parental Education: Did not finish High School	0.08	0.28	0.00	1.00
Parental Education: Graduated High School	0.18	0.39	0.00	1.00
Parental Education: Some education after High School	0.20	0.40	0.00	1.00
Parental Education: Graduated College	0.53	0.50	0.00	1.00
Computer at Home	0.90	0.30	0.00	1.00
Books at Home: 0-10	0.13	0.34	0.00	1.00
Books at Home: 11-25	0.22	0.41	0.00	1.00
Books at Home: 26-100	0.35	0.48	0.00	1.00
Books at Home: >100	0.30	0.46	0.00	1.00

Notes: Continuation on next page

Descriptive statistics (continued)

	Mean	Std. Dev.	Min.	Max.
Instructional Focus Outcomes:				
Instruction Time	0.41	0.49	0.00	1.00
Instructional Resources	0.60	0.49	0.00	1.00
Differentiated Instruction: Standards	0.44	0.50	0.00	1.00
Differentiated Instruction: Material	0.65	0.48	0.00	1.00
Differentiated Instruction: Activities	0.41	0.49	0.00	1.00
Differentiated Instruction: Methods	0.62	0.49	0.00	1.00
Differentiated Instruction: Pace	0.59	0.49	0.00	1.00
Teacher dedication: Discuss students' performance	0.56	0.50	0.00	1.00
Teacher dedication: Set goals	0.40	0.49	0.00	1.00
Teacher dedication: Discuss goals	0.41	0.49	0.00	1.00
Teacher dedication: Adjust teaching	0.64	0.48	0.00	1.00
Teacher Controls:				
Teacher Race/Ethnicity: White	0.83	0.38	0.00	1.00
Teacher Race/Ethnicity: Black	0.07	0.25	0.00	1.00
Teacher Race/Ethnicity: Hispanic	0.06	0.24	0.00	1.00
Teacher Race/Ethnicity: Asian	0.02	0.15	0.00	1.00
Teacher Race/Ethnicity: Other	0.01	0.11	0.00	1.00
Teacher Education: Bachelor or less	0.50	0.50	0.00	1.00
Teacher Education: Master or more	0.50	0.50	0.00	1.00
NBPTS Teacher Certificate: Yes	0.13	0.34	0.00	1.00
NBPTS Teacher Certificate: Working towards	0.02	0.15	0.00	1.00
NBPTS Teacher Certificate: No	0.85	0.36	0.00	1.00
Alternative Teacher Certificate: Yes	0.13	0.33	0.00	1.00
Alternative Teacher Certificate: No	0.87	0.33	0.00	1.00
Teacher Experience: 2 years or less	0.09	0.28	0.00	1.00
Teacher Experience: 3-5 years	0.14	0.34	0.00	1.00
Teacher Experience: 6-10 years	0.22	0.42	0.00	1.00
Teacher Experience: 11-20 years	0.27	0.44	0.00	1.00
Teacher Experience: 21 years or more	0.28	0.45	0.00	1.00

Notes: Descriptive statistics (mean, standard deviation, minimum, maximum) for main treatment, outcome, and control variables. NBPTS stands for National Board for Professional Teaching Standards. Data source: See Figure 4.2

Table A4.3: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by subjects

	(1) Science	(2) Civics	(3) Economics	(4) Geography	(5) History
CCSS Exposure	-0.096**	0.010	-0.052	-0.036	-0.004
	(0.042)	(0.087)	(0.307)	(0.079)	(0.097)
State and Year FEs	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Adj. R-squared	0.402	0.391	0.381	0.429	0.380
Observations	931,600	55,150	19,930	32,130	64,810

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subject indicated in the column header. Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.4: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by grades

	(1) Grade 4	(2) Grade 8	(3) Grade 12
CCSS Exposure	-0.134***	-0.007	-0.005
	(0.036)	(0.062)	(0.072)
State and Year FEs	YES	YES	YES
Controls	YES	YES	YES
Adj. R-squared	0.376	0.433	0.370
Observations	434,440	582,590	86,600

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.5: Effect of CCSS exposure on student achievement in non-targeted subjects, subgroups by quartiles of states' student achievement before 2010

	(1) Quartile 1 (lowest)	(2) Quartile 2	(3) Quartile 3	(4) Quartile 4 (highest)
CCSS Exposure	0.004	-0.097*	-0.061	-0.041
	(0.046)	(0.048)	(0.051)	(0.029)
State and Year FEs	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Adj. R-squared	0.413	0.365	0.353	0.380
Observations	315,160	334,740	233,430	220,300

Notes: Each entry is from a separate two-way fixed effects regression model. Sample of Quartile 1 subgroup includes students from states in the lowest quartile with respect to average student achievement in years before 2010. Sample of Quartile 2 subgroup includes students from states in the second lowest quartile with respect to average student achievement in years before 2010. Quartile 3 and 4 defined accordingly. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.6: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for public schooling policies

	Control for:							
	(1) NCLB/ESSA Waivers	(2) State Proficiency	(3) NGSS Adoption	(4) Expenditures	(5) Teacher Policies	(6) School Choice	(7) Evolution	(8) Charter Schools
CCSS Exposure	-0.074**	-0.083**	-0.066*	-0.077**	-0.098**	-0.075**	-0.079**	-0.080**
	(0.035)	(0.037)	(0.034)	(0.035)	(0.037)	(0.037)	(0.036)	(0.036)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.390	0.390	0.390	0.391	0.397	0.390	0.390	0.390
Observations	1,103,630	1,089,080	1,103,630	1,103,620	769,410	1,099,880	1,099,880	1,099,880

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Additional policy controls (state-by-year level, unless otherwise stated): Model 1 controls for NCLB/ESSA requirements waiver; Model 2 controls for percentile ranking of state's proficiency standards (average over math and ELA proficiency for grades 4 and 8; missing years are imputed as average of adjacent years); Model 3 controls for adoption of Next Generation Science Standards or of standards based on Next Generation Science Standards framework; Model 4 controls for district-by-year-level per-pupil education expenditures in logarithmized dollars; Model 5 controls for index of teacher quality policies; Model 6 controls for public school choice laws; Model 7 controls for laws permitting public school teachers to teach 'weaknesses of evolution'; Model 8 controls for charter school laws. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: National Center for Education Statistics (Local Education Agency (School District) Finance Survey F-33); Sorens et al. (2008); Ross et al. (2017); Hamlin and Peterson (2018); Jordan and Grossmann (2020); See Figure 4.2

Table A4.7: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for private schooling policies

	Control for:					
	(1) State Approval	(2) Licensure of Teachers	(3) Registration	(4) Curriculum	(5) Tax Credits	(6) Vouchers
CCSS Exposure	-0.079**	-0.080**	-0.075**	-0.080**	-0.088***	-0.064*
	(0.036)	(0.036)	(0.036)	(0.036)	(0.028)	(0.035)
State and Year FEs	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.390	0.390	0.390	0.390	0.390	0.390
Observations	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Additional policy controls (state-by-year level, unless otherwise stated): Model 1 controls for mandatory state approval, where state has discretion, licensing, or accreditation of private schools; Model 2 controls for mandatory state licensure of private school teachers; Model 3 controls for mandatory registration or licensing of private schools (note: if approval is required, registration is also coded as being required); Model 4 controls for extent of private school curriculum control; Model 5 controls for tax credit/deduction law for scholarship contributions or educational expenses of parents; Model 6 controls for publicly funded voucher laws. Standard errors clustered at the state level. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: Sorens et al. (2008); Jordan and Grossmann (2020); See Figure 4.2

Table A4.8: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for homeschooling policies

	Control for:							
	(1) Curriculum	(2) Statute	(3) Notice Extent	(4) Notice Frequency	(5) Notice Index	(6) Recordkeeping	(7) Testing	(8) Teachers
CCSS Exposure	-0.082**	-0.080**	-0.079**	-0.086**	-0.089**	-0.080**	-0.079**	-0.080**
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.035)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.390	0.390	0.390	0.390	0.391	0.390	0.390	0.390
Observations	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880	1,099,880

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Additional policy controls (state-by-year level, unless otherwise stated): Model 1 controls for subjects/curriculum requirement for homeschoolers; Model 2 controls for whether homeschooling is explicitly permitted by statute; Model 3 controls for extent of homeschooling notice requirement; Model 4 controls for frequency of homeschooling notice requirement; Model 5 controls for homeschooling notification index (Extent of homeschooling notice requirement * Frequency of homeschooling notice requirement); Model 6 controls for extent of homeschool recordkeeping requirements; Model 7 controls for standardized testing or other official evaluation requirement of homeschooling; Model 8 controls for homeschooling teacher qualifications requirement. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: Sorens et al. (2008); Jordan and Grossmann (2020); See Figure 4.2

Table A4.9: Effect of CCSS exposure on student achievement in non-targeted subjects, robustness with additional controls for compulsory schooling policies

	Control for:			
	(1) Compulsory school age, lower bound	(2) Compulsory school age, upper bound	(3) Compulsory school years	(4) Kindergarten attendance
CCSS Exposure	-0.079**	-0.079**	-0.081**	-0.082**
	(0.037)	(0.035)	(0.036)	(0.036)
State and Year FEs	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Adj. R-squared	0.390	0.390	0.390	0.390
Observations	1,099,880	1,099,880	1,099,880	1,099,880

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Additional policy controls (state-by-year level, unless otherwise stated): Model 1 controls for compulsory school age, lower bound (minimum standard if set by local school district; age at which parental waivers not permitted); Model 2 controls for compulsory school age, upper bound (minimum standard if set by local school district; age at which parental waivers not permitted); Model 3 controls for compulsory school years (Compulsory school age, upper bound – Compulsory school age, lower bound); Model 4 controls for kindergarten attendance requirement. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: Sorens et al. (2008); Jordan and Grossmann (2020); See Figure 4.2

Table A4.10: Effect of CCSS exposure on student achievement in non-targeted subjects, further robustness checks

	(1) Exclude already-treated states from controls	(2) Controls: No parental education	(3) Controls: No parental education & No imputation of missings	(4) Treatment: Number of years of CCSS Exposure	(5) Sample: No Charter Schools	(6) No weights
CCSS Exposure	-0.082**	-0.079**	-0.090**	-0.006	-0.079**	-0.093***
	(0.036)	(0.036)	(0.036)	(0.007)	(0.036)	(0.025)
State and Year FEs	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.392	0.380	0.381	0.390	0.390	0.399
Observations	996,390	1,103,630	1,067,950	1,103,630	1,077,420	1,103,630

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects not targeted by the CCSS (Pool of science, civics, economics, geography, and history). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Model 1 excludes states which adopted the CCSS in 2011 and 2012 from the sample which implies that no students from already-treated states act as controls; Model 2 excludes parental education from set of control variables; Model 3 excludes parental education from set of control variables and does not impute other missing control variables; Model 4 defines the explanatory variable as the number of schooling years a student was exposed to CCSS (at the time of testing); Model 5 excludes charter schools from the sample of public schools; Model 6 does not use population weights. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.11: Effect of CCSS exposure on instruction time in non-targeted subjects, subgroups by teacher characteristics

	Teacher Race/Ethnicity					Teacher education		Teacher certification (NBPTS)			Teacher experience in years				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	White	Black	Hispanic	Asian	Other	Bachelor or less	Master or more	Yes	Working towards	No	2 or less	3-5	6-10	11-20	21 or more
CCSS Exposure	-0.201***	-0.050	0.033	-0.081	0.049	-0.145**	-0.226***	-0.147**	0.120	-0.202***	-0.170*	-0.074	-0.207***	-0.004	-0.351***
	(0.048)	(0.083)	(0.052)	(0.169)	(0.182)	(0.055)	(0.064)	(0.067)	(0.124)	(0.061)	(0.088)	(0.057)	(0.067)	(0.075)	(0.088)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.093	0.164	0.134	0.288	0.365	0.089	0.120	0.153	0.310	0.106	0.146	0.138	0.127	0.123	0.132
Observations	696,350	63,680	39,040	24,130	15,010	404,030	443,810	77,450	16,460	502,330	88,620	128,950	190,000	247,230	192,820

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Probability that students' teacher reports about subject in question that students receive more than five hours of weekly instruction. Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for teacher race/ethnicity, teacher education, teacher certification (separate indicator variables for certification of National Board for Professional Teaching Standards, and for alternative certification), and teacher experience, as well as student controls for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.12: Effect of CCSS exposure on instructional resources in non-targeted subjects, subgroups by teacher characteristics

	Teacher Race/Ethnicity					Teacher education		Teacher certification (NBPTS)			Teacher experience in years				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	White	Black	Hispanic	Asian	Other	Bachelor or less	Master or more	Yes	Working towards	No	2 or less	3-5	6-10	11-20	21 or more
CCSS Exposure	-0.121***	-0.038	-0.220***	-0.071	0.177***	-0.121***	-0.103***	-0.132***	-0.163*	-0.122***	-0.127**	-0.054**	-0.148***	-0.110**	-0.115***
	(0.028)	(0.047)	(0.028)	(0.091)	(0.061)	(0.030)	(0.031)	(0.033)	(0.096)	(0.039)	(0.056)	(0.026)	(0.029)	(0.045)	(0.028)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.044	0.051	0.046	0.078	0.095	0.038	0.050	0.056	0.080	0.039	0.049	0.038	0.047	0.049	0.058
Observations	647,100	58,340	34,910	23,110	13,460	373,640	412,020	71,420	15,520	464,110	82,790	120,210	174,310	229,130	179,120

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Probability that students' teacher reports about subject in question that their school system provides them with all or most materials and other resources they need for the instruction. Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for teacher race/ethnicity, teacher education, teacher certification (separate indicator variables for certification of National Board for Professional Teaching Standards, and for alternative certification), and teacher experience, as well as student controls for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.13: Effect of CCSS exposure on the quality of teacher-student interactions (setting goals) in non-targeted subjects, subgroups by teacher characteristics

	Teacher Race/Ethnicity					Teacher education		Teacher certification (NBPTS)			Teacher experience in years				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	White	Black	Hispanic	Asian	Other	Bachelor or less	Master or more	Yes	Working towards	No	2 or less	3-5	6-10	11-20	21 or more
CCSS Exposure	-0.029*	-0.064	-0.026	0.009	-0.195	-0.030	-0.046**	0.010	-0.079	-0.047**	-0.054	-0.012	-0.050	-0.022	-0.024
	(0.016)	(0.048)	(0.038)	(0.170)	(0.210)	(0.022)	(0.023)	(0.037)	(0.085)	(0.020)	(0.052)	(0.028)	(0.044)	(0.017)	(0.038)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.059	0.071	0.149	0.106	0.101	0.099	0.086	0.117	0.140	0.078	0.102	0.101	0.094	0.083	0.095
Observations	458,530	39,930	24,280	16,870	10,040	250,090	304,390	71,120	15,490	462,530	57,820	81,750	123,090	170,430	121,310

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Probability that students' teacher reports about subject in question that she sets goals for specific progress the student would like to make at least once a month. Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for teacher race/ethnicity, teacher education, teacher certification (separate indicator variables for certification of National Board for Professional Teaching Standards, and for alternative certification), and teacher experience, as well as student controls for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.14: Effect of CCSS exposure on the quality of teacher-student interactions (discussing goals) in non-targeted subjects, subgroups by teacher characteristics

	Teacher Race/Ethnicity					Teacher education		Teacher certification (NBPTS)			Teacher experience in years				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	White	Black	Hispanic	Asian	Other	Bachelor or less	Master or more	Yes	Working towards	No	2 or less	3-5	6-10	11-20	21 or more
CCSS Exposure	-0.033	-0.086	-0.054	-0.042	-0.251	-0.053	-0.042**	0.007	-0.123	-0.059**	-0.070	-0.052	-0.029	-0.046**	-0.038
	(0.020)	(0.052)	(0.036)	(0.167)	(0.200)	(0.033)	(0.017)	(0.034)	(0.079)	(0.027)	(0.069)	(0.043)	(0.030)	(0.022)	(0.040)
State and Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.060	0.076	0.127	0.116	0.116	0.094	0.086	0.113	0.140	0.077	0.094	0.104	0.094	0.080	0.096
Observations	458,460	39,940	24,270	16,870	10,020	250,020	304,350	71,120	15,480	462,400	57,780	81,770	123,080	170,360	121,290

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Probability that students' teacher reports about subject in question that she discusses progress her student has made toward goals previously set at least once a month. Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for teacher race/ethnicity, teacher education, teacher certification (separate indicator variables for certification of National Board for Professional Teaching Standards, and for alternative certification), and teacher experience, as well as student controls for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

A4.2 Analysis of the Effects of the CCSS on targeted subjects

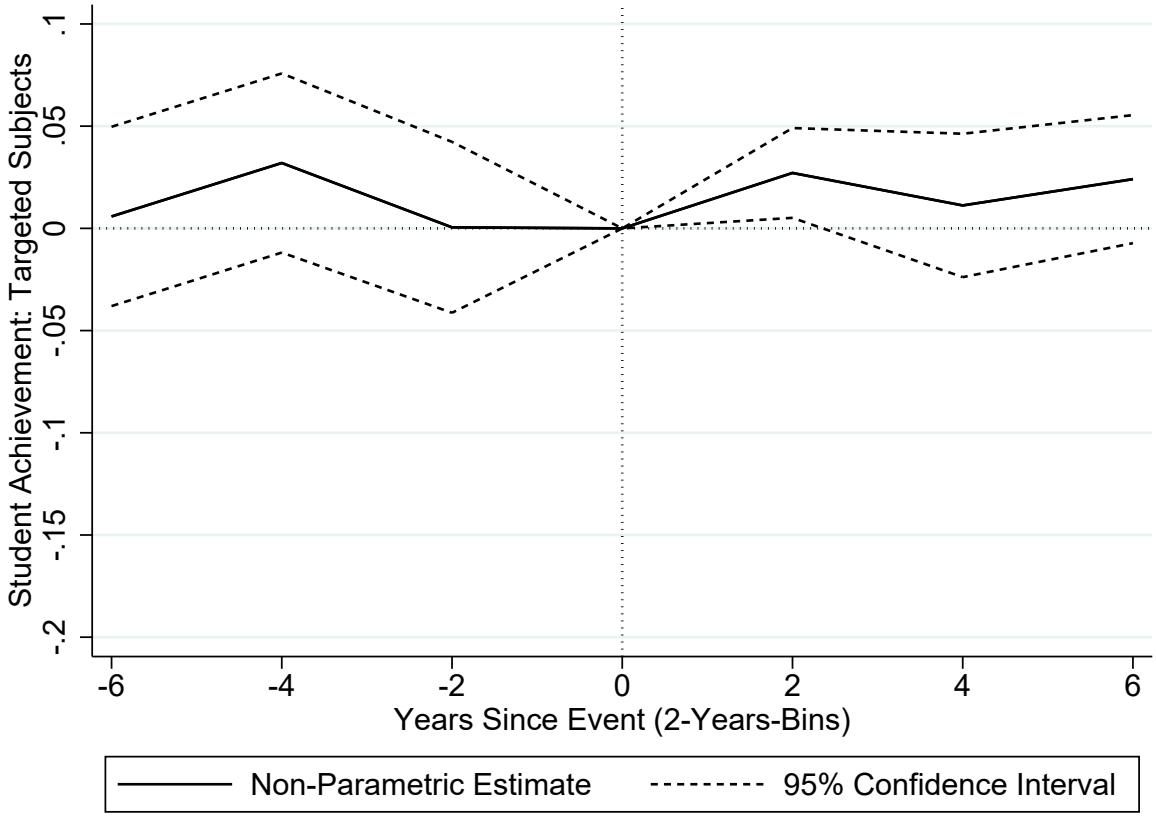
We show evidence that the CCSS had, at best, modestly positive effects on student achievement in the targeted subjects math and ELA. This analysis largely confirms the conclusions Bleiberg (2021) has drawn on this question, although our findings (using more data, among other conceptual differences) suggest that Bleiberg (2021) rather overestimates than underestimates the positive effects of the CCSS on student achievement in targeted subjects.

First, we visualize the modest positive effects in an event-study graph depicted in Figure A4.1. The estimation equation follows equation (4.2) presented in Section 4.3, with T_{istuv} now pooling standardized student achievement in all subjects that are targeted by the CCSS, across all available grades.[1] The event-study graph shows a modest increase in student achievement after the adoption of the CCSS that is marginally significant directly after adoption and insignificantly different from zero thereafter. Using a second dataset on student achievement from the Stanford Education Data Archive SEDA 4.0 (Reardon et al., 2021) with a shorter pre- but longer post-period relative to the NAEP, we find basically null effects, see Figure A4.2. SEDA does not contain data on non-targeted subjects which is why we cannot use it for the main analysis.

Second, we also show parametric two-way fixed effects DD results. The estimation equation follows equation (4.1) presented in Section 4.3, with T_{istuv} pooling standardized student achievement in all subjects that are targeted by the CCSS, across all available grades (as above). As reported in Table A4.15, the baseline specification and the econometric robustness specifications including state-specific time trend controls and the DDD model yield zero to small positive point estimates. Next, we analyse robustness regarding the treatment indicator as described in Section 4.6 and Appendix A4.3. As shown in Table A4.16, we find zero to modestly positive effects across specifications.

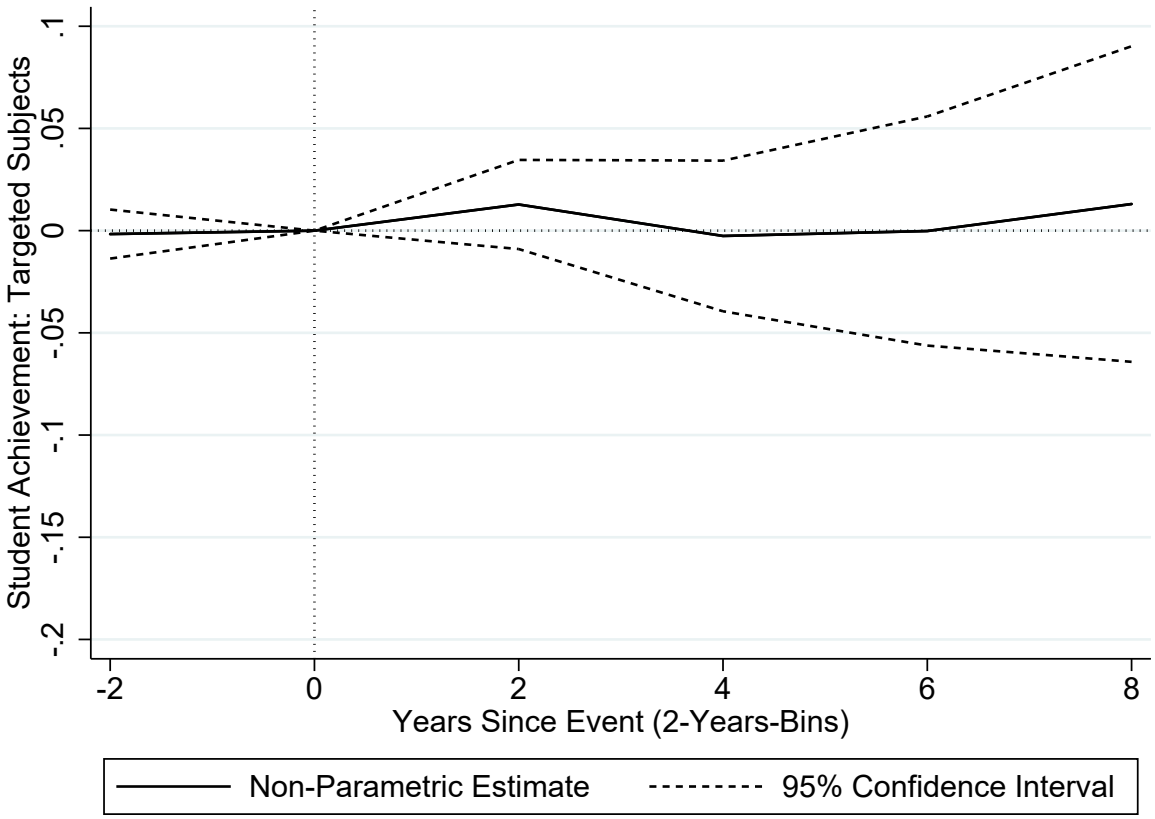
[1] Table A4.17 provides the list of grades in which the NAEP tests were administered for each of the four targeted subjects math, reading, writing, and vocabulary (and for which state identifiers of students are available). We use the student-level data from all these subject-year-grade combinations in our analysis.

Figure A4.1: Event-study graph: Targeted subjects (NAEP)



Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Standardized student achievement in subjects targeted by the CCSS (Pool of math, reading, vocabulary and writing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as state, test year, grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Numbers on horizontal axis refer to respective two-year bins; i.e. 2 = first two years of treatment (year 0 = excluded category). The p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.003 and 0.075, respectively. Data source: See Figure 4.2

Figure A4.2: Event-study graph: Targeted subjects (SEDA)



Note: Coefficients from non-parametric event-study regressions and their 95% confidence intervals. Dependent variable: Standardized student achievement in subjects targeted by the CCSS (Pool of math and ELA). Controls: District shares of races/ethnicities, English language learner status, disability status, subsidized lunch status, economic disadvantage, rural location, as well as state, test year, grade and subject fixed effects. Regressions use precision weights (the inverse of the standard error of average student achievement in math and ELA squared) and standard errors clustered at the state level. Numbers on horizontal axis refer to respective two-year bins; i.e. 2 = first two years of treatment (year 0 = excluded category). The p values of omnibus hypothesis tests of zero pre- and post-event effects are 0.782 and 0.190, respectively. Data source: Reardon et al. (2021)

Table A4.15: Effect of CCSS exposure on student achievement in targeted subjects, econometric robustness

	(1) DD	(2) (1) + linear state trends	(3) (2) + quadratic state trends	(4) (3) + cubic state trends	(5) DDD	(6) DD with private schools
CCSS Exposure	0.010	0.004	0.002	-0.001	0.007	0.011
	(0.023)	(0.029)	(0.051)	(0.051)	(0.055)	(0.020)
State and Year FEs	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.352	0.353	0.354	0.354	0.353	0.349
Observations	6,392,940	6,392,940	6,392,940	6,392,940	6,550,510	6,550,510

Notes: Each entry is from a separate two-way fixed effects regression model, where Model (1) is the baseline model, Models (2), (3), (4) subsequently add linear, quadratic, and cubic state-specific time trends, and Model (5) presents a triple-difference model where school type (public vs. private school students), school type*state, school type*year, and state*year fixed effects replace all state-specific time trends. Model (6) estimates the basic two-way fixed effects model on a sample of public and private school students. Dependent variable: Standardized student achievement in subjects targeted by the CCSS (Pool of math, reading, vocabulary and writing). Explanatory variable: Share of schooling years a student was exposed to CCSS (at the time of testing). Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.16: Effect of CCSS exposure on student achievement in targeted subjects using different definitions of treatment implementation

	CCSS adoption	CCSS implementation requirement	CCSS implementation strategies	Effective CCSS implementation	Include temporary CCSS adopters and implementers	CCSS-aligned testing
	(1)	(2)	(3)	(4)	(5)	(6)
CCSS Exposure	0.010	-0.002	0.019	-0.001	0.046***	-0.007
	(0.023)	(0.027)	(0.019)	(0.021)	(0.017)	(0.032)
State and Year FEs	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.352	0.352	0.352	0.352	0.352	0.352
Observations	6,392,940	6,392,940	6,392,940	6,392,940	6,392,940	6,392,940

Notes: Each entry is from a separate two-way fixed effects regression model. Dependent variable: Standardized student achievement in subjects targeted by the CCSS (Pool of math, reading, vocabulary and writing). Explanatory variables: Share of schooling years a student was exposed to CCSS (at the time of testing), where in Models 1 (CCSS adoption, baseline model) each schooling year counts as exposed in a given state in which the state adopted the CCSS permanently before that year or in the same year according to Achieve Inc. (2013) and CCSSI (2022), where Model 2 (CCSS implementation requirement) each schooling year counts as exposed in a given state in which the state expects teachers to fully incorporate CCSS into classroom instruction in grades K-12 in English language arts and mathematics according to Achieve Inc. (2013) and CCSSI (2022), where in Models 3 (CCSS implementation strategies) each schooling year counts as exposed in a given state if state education agency officials report that their state pursued at least two out of three CCSS implementation strategies (professional development, new instructional materials, joined testing consortium) as reported in Webber et al. (2014), where in Models 4 (Effective CCSS implementation) each schooling year counts as exposed in a given state in which the state implemented an effective change in state standard content through the adoption of CCSS which we define as not having had a state standard in place before the adoption of CCSS whose academic rigor is "too close to call" in comparison with CCSS (Carmichael et al., 2010) for the set of states adopting CCSS according to Achieve Inc. (2013) and CCSSI (2022), where in Models 5 (Include temporary CCSS adopters) each schooling year counts as exposed in a given state in which a state adopted and/or implemented CCSS at least temporarily according to Bleiberg (2021); and where in Models 6 (CCSS-aligned testing) each schooling year counts as exposed in a given state in which the state adopted CCSS-aligned standardized testing including field and transitional tests according to our own research (see Table A4.18 for state-specific details of CCSS-aligned testing). Table A4.19 provides state-specific coding information on all treatment definitions. Controls: Indicator variables for gender, races/ethnicities, English language learner status, disability status, subsidized lunch status, parental education, home possessions (separate indicator variables for computer and books) as well as grade and subject fixed effects. Regressions use population weights and standard errors clustered at the state level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: See Figure 4.2

Table A4.17: List of grades of NAEP tests for targeted subjects

	Targeted Subjects			
Year	Math	Reading	Vocabulary	Writing
2002		4, 8, 12		8, 12
2003	4, 8	4, 8		
2004				
2005	4, 8, 12	4, 8, 12		4, 8, 12
2006				
2007	4, 8	4, 8		8, 12
2008				
2009	4, 8, 12	4, 8, 12	4, 8, 12	
2010				
2011	4, 8	4, 8	4, 8	8, 12
2012				
2013	4, 8	4, 8		
2014				
2015	4, 8, 12	4, 8, 12		
2016				
2017	4, 8,	4, 8		

Notes: NAEP student achievement data in targeted subjects at the subject-by-year-by-grade level. Data source: See Figure 4.2

A4.3 Background Information on Treatment Definition Robustness

To test whether our results hold if we define treatment based on CCSS implementation, we re-run our main regression using five different treatment variables each capturing different information about CCSS implementation in Section 4.6. This appendix provides background information about the construction and data sources of these five alternative treatment definitions.

First, we collect information on CCSS implementation requirements, from Achieve Inc. (2013) and CCSSI (2022). Here, the year of full implementation of CCSS is defined as the school year the respective state expects teachers in grades K-12 in math and ELA to incorporate the standards into classroom instruction. The time between adoption and full implementation varies between 1 to 4 years across adopting states, with an average of about 3 years.

Second, we note that state expectations about teachers implementing the CCSS into classroom instruction do not necessarily have to be aligned with actual state efforts to implement the CCSS. However, the latter might be more relevant for ultimate exposure of students to the CCSS and potential effects on student achievement than formal state expectations. To incorporate this idea into our analysis, we make use of a survey of state education agency officials provided by Webber et al. (2014). They conducted a survey of state education agency officials which collects information on actual state efforts towards CCSS implementation. Specifically, the survey respondents answer questions about whether the state has provided, guided or funded professional development on the CCSS, whether it has provided curriculum or instructional materials for the CCSS, and whether it has worked with a federally funded consortium to develop assessments aligned with the CCSS. In this treatment coding, we count a schooling year in a given state as being exposed to the CCSS if this state has adopted the CCSS according to Achieve Inc. (2013) and CCSSI (2022) and pursued at least two out of three CCSS implementation strategies as reported by the state education agency officials.

Third, we calculate a treatment indicator capturing effective CCSS implementation. Here, we build on the idea that effective change of the state standard can only be induced by the CCSS if the state standard in place prior to the adoption of the CCSS in the state in question is sufficiently different from the CCSS. To this end, we make use of a comparison of academic rigor of the CCSS with the respective state standards in place prior to the CCSS provided by Carmichael et al. (2010). We code students from states as being in the control group at all years if their pre-CCSS state standards are “too close to call” in both math and ELA in a comparison with the CCSS (in addition to coding students from states that did not adopt the CCSS according to Achieve Inc. (2013) and CCSSI (2022) as being in the control group at all years).

Fourth, we account for the fact that some states may have adopted and/or implemented some elements of the CCSS temporarily, even when they are listed as non-adopters and non-implementers in Achieve Inc. (2013) and CCSSI (2022). According to Bleiberg (2021), four of

the eight non-permanent adopters of the CCSS in the coding based on Achieve Inc. (2013) and CCSSI (2022) have implemented at least some elements of the CCSS temporarily. The map presented in Figure 4.1 depicts them as temporary adopters. In this treatment coding, we count a schooling year as being exposed to the CCSS if the state in question adopted the CCSS temporarily or permanently.

Fifth, we argue that the relevant criterion for actual CCSS implementation might be the alignment of the content of state-mandated standardized testing with the CCSS. To assess this hypothesis, we did own background research to find out which state mandated what type of standardized test for each grade group and year. State-specific details on which tests (including field and transitional tests) are mandated when and for which grade are reported in Table A4.18. Subsequently, we assessed which of these tests are aligned with the CCSS. This analysis allowed us to infer the year in which CCSS-aligned standardized testing was mandated in a given state. In the corresponding treatment coding, we count a schooling year in a given state as being exposed to the CCSS if this state has mandated CCSS-aligned standardized testing in any group grade in that year.

Table A4.19 presents the treatment status for each state for the baseline definition of CCSS adoption and the five definitions of CCSS implementation. In particular, it shows whether schooling years in a given state never count as being exposed to the CCSS ("always control"), or, if they do, from which year onwards.

Table A4.18: State-mandated tests, by state and grade group from 2010 onwards (based on own research)

State	3-8 grades	High school
Alabama	**2010**: Alabama Reading and Math Test (ARMT+) **2014**: ACT Aspire	**2010**: Alabama High School Graduation Exam (AHSGE) **2014**: ACT End of course
Alaska	**2010**: Standards-Based Assessments (SBAs) **2015**: Alaska Measures of Progress (AMP) **2017**: Performance Evaluation for Alaska's Schools (PEAKS)	**2010**: SBAs **2015**: AMP; ACT, SAT, or WorkKeys **2017**: PEAKS; ACT, SAT, or WorkKeys
Arizona	**2010**: Arizona Instrument to Measure Standards (AIMS) **2014**: Field test PARCC **2015**: AzMerit	**2010**: AIMS **2014**: Field test PARCC **2015**: AzMerit
Arkansas	**2010**: Arkansas Comprehensive Testing, Assessment, and Accountability Program (ACTAAP) **2013**: Arkansas Benchmark **2014**: Field test PARCC **2015**: PARCC **2016**: ACT Aspire	**2010**: Arkansas Comprehensive Testing, Assessment, and Accountability Program (ACTAAP) **2013**: Arkansas Benchmark **2014**: Field test PARCC **2015**: PARCC **2016**: ACT Aspire

Table A4.18 - Aligned Testing Details (continued)

State	3-8 grades	High school
California	**2010**: Standardized Testing and Reporting (STAR) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Standardized Testing and Reporting (STAR) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Colorado	**2010**: Colorado Student Assessment Program (CSAP) **2012**: Transitional Colorado Assessment Program (TCAP) **2014**: Field test PARCC **2015**: PARCC	**2010**: Colorado Student Assessment Program (CSAP) **2012**: Transitional Colorado Assessment Program (TCAP) **2014**: Field test PARCC **2015**: PARCC **2016**: PSAT, ACT **2017**: SAT
Connecticut	**2010**: Connecticut Mastery Test (CMT) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Connecticut Academic Performance Test (CAPT) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: SAT
Delaware	**2010**: Delaware Comprehensive Assessment System (DCAS) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Delaware Comprehensive Assessment System (DCAS) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: SAT
District of Columbia	**2010**: District of Columbia Comprehensive Assessment System (DC CAS) **2012**: DC CAS revised (transitional test) **2013**: DC CAS revised **2014**: Field test PARCC **2015**: PARCC	**2010**: District of Columbia Comprehensive Assessment System (DC CAS) **2012**: DC CAS revised (transitional test) **2013**: DC CAS revised **2014**: Field test PARCC **2015**: PARCC
Florida	**2010**: Florida Comprehensive Assessment Test (FCAT) **2011**: FCAT 2.0 **2014**: Florida Standards Assessment (FSA)	**2010**: FCAT **2011**: Florida End-of-Course (EOC) Assessments **2014**: FSA or Next Generation Sunshine State Standards (NGSSS) **2016**: FSA
Georgia	**2010**: Criterion-Referenced Competency Tests (CRCT) **2015**: Georgia Milestones Assessment System (GMAS)	**2010**: End of Course Test (EOCT) **2015**: GMAS
Hawaii	**2010**: Hawaii State Assessment (HSA) **2014**: Part-HAS Part-Smarter Balanced (transition test) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Hawaii State Assessment (HSA) **2014**: Part-HAS Part-Smarter Balanced (transition test) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced

Table A4.18 - Aligned Testing Details (continued)

State	3-8 grades	High school
Idaho	**2010**: Idaho Standards Achievement Test (ISAT) **2013**-14: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Idaho Standards Achievement Test (ISAT) **2013**-14: Field test Smarter Balanced **2015**: Smarter Balanced in 10th grade; Choice of ACT, SAT or ACT Compass for 11th grade.
Illinois	**2010**: Illinois Standards Achievement Tests (ISAT) **2014**: Field test PARCC **2015**: PARCC	**2014**: Field test PARCC **2015**: PARCC **2016**: SAT
Indiana	**2010**: Indiana Statewide Testing for Educational Progress Plus (ISTEP+)	**2010**: ISTEP+, end-of-course tests
Iowa	**2010**: Iowa Test of Basic Skills (ITBS) **2011**: Iowa Assessments **2014**: Field test Smarter Balanced	**2010**: Iowa Test of Educational Development (ITED) **2011**: Iowa Assessments **2014**: Field test Smarter Balanced
Kansas	**2014**: No test **2015**: Field test Kansas State Assessment (KSA) **2016**: KSA	**2014**: No test **2015**: Field test Kansas State Assessment (KSA) **2016**: KSA
Kentucky	**2010**: Kentucky Performance Rating for Educational Progress (K-PREP)	**2010**: K-PREP, ACT QualityCore, ACT
Louisiana	**2006**: Louisiana Educational Assessment Program (LEAP) and iLEAP **2013**: LEAP and iLEAP revised (transitional test) **2014**: Field test PARCC **2015**: PARCC **2016**: Mix of PARCC and LEAP	**2010**: End-of-course tests, ACT, ACT Plan **2013**-14: End-of-course revised (transitional test) **2015**: End-of-course revised
Maine	**2010**: New England Common Assessment Program (NECAP) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: Maine Educational Assessments (MEA)	**2010**: SAT **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2015**: SAT
Maryland	**2010**: Maryland State Assessment (MSA) **2014**: Field test PARCC **2015**: PARCC	**2010**: Maryland High School Assessment (HSA) **2014**: Field test PARCC **2015**: PARCC
Massachusetts	**2010**: Massachusetts Comprehensive Assessment System (MCAS) **2014**: Field test PARCC **2014**: Districts choose between PARCC or MCAS **2016**: Mix of PARCC and Next Generation MCAS **2017**: Next Generation MCAS	**2010**: Massachusetts Comprehensive Assessment System (MCAS) **2014**: Field test PARCC **2014**: Districts choose between PARCC or MCAS **2015**: MCAS

Table A4.18 - Aligned Testing Details (continued)

State	3-8 grades	High school
Michigan	**2010**: Michigan Educational Assessment Program (MEAP) **2014**: Field test Smarter Balanced **2015**: Michigan Student Test of Educational Progress (M-STEP)	**2010**: Michigan Merit Exam (MME: includes SAT, WorkKeys) **2015**: MME, PSAT
Minnesota	**2010**: Minnesota Comprehensive Assessments (MCA)	**2010**: MCA
Mississippi	**2010**: Mississippi Curriculum Test (MCT) **2014**: Field test PARCC **2015**: PARCC **2016**: Mississippi Academic Assessment Program (MAAP)	**2010**: Subject Area Testing Program (SATP) **2014**: Field test PARCC **2015**: PARCC **2016**: ACT
Missouri	**2010**: Missouri Assessment Program (MAP) **2014**: MAP revised (transitional test) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: MAP	**2010**: Missouri End-of-Course Assessments **2014**: End of Course (EOC) revised (transitional test) **2015**: EOC revised, ACT
Montana	**2010**: Montana's Criterion Reference Test (Montana's CRT) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Montana's CRT **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: ACT
Nebraska	**2010**: Nebraska State Accountability Tests (NeSA)	**2010**: NeSA **2017**: ACT
Nevada	**2010**: Nevada's Criterion Reference Test (Nevada's CRT) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: High School Proficiency Examination (HSPE) **2015**: ACT
New Hampshire	**2010**: New England Common Assessment Program (NECAP) **2013**: NECAP revised (transitional test) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: NECAP **2013**: NECAP revised (transitional test) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: PACE, SAT
New Jersey	**2010**: New Jersey Assessment of Skills and Knowledge (NJASK) **2014**: NJASK revised (transitional test) **2014**: Field test PARCC **2015**: PARCC	**2010**: High School Proficiency Assessment **2014**: Field test PARCC **2015**: PARCC
New Mexico	**2010**: New Mexico Standards-based Assessment (NMSBA) **2014**: Field test PARCC **2015**: PARCC	**2010**: NMSBA **2014**: Field test PARCC **2015**: PARCC
New York	**2012**: Field test **2013**: New York State English Language Arts and Mathematics Tests **2014**: Field test PARCC **2016**: New York State Assessments	**2013**: Regents Exams **2014**: Regents revised **2014**: Field test PARCC

Table A4.18 - Aligned Testing Details (continued)

State	3-8 grades	High school
North Carolina	**2012**: Field test **2013**: End-of-grade tests **2014**: Field test Smarter Balanced	**2012**: Field test **2013**: End-of-course tests, ACT PLAN, ACT, WorkKeys **2014**: Field test Smarter Balanced
North Dakota	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Ohio	**2014**: Ohio Achievement Assessments **2014**: Field Test PARCC **2015**: PARCC **2016**: Ohio State Tests (OST)	**2014**: Ohio Graduation Tests **2014**: Field Test PARCC **2015**: PARCC **2016**: OST, Ohio Graduation Test **2017**: End-of-course tests, SAT/ACT
Oklahoma	**2010**: Oklahoma Core Curriculum Test (OCCT) **2017**: Oklahoma School Testing Program (OSTP)	**2010**: End-of-course tests **2017**: OSTP
Oregon	**2010**: Oregon Assessment of Knowledge and Skills **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: Oregon Assessment of Knowledge and Skills **2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Pennsylvania	**2010**: Pennsylvania System of School Assessment (PSSA) **2013**: Field test PSSA revised **2015**: PSSA revised	**2010**: Pennsylvania System of School Assessment (PSSA) **2013**: Keystone Exams
Rhode Island	**2010**: New England Common Assessment Program (NECAP) **2014**: Field test PARCC **2015**: PARCC	**2010**: New England Common Assessment Program (NECAP) **2014**: Field test PARCC **2015**: PARCC
South Carolina	**2010**: South Carolina Palmetto Assessment of State Standards (SCPASS) **2015**: ACT Aspire **2016**: SC Ready	**2014**: ACT Plus Writing, ACT WorkKeys **2015**: End-of-course tests, ACT
South Dakota	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Tennessee	**2014**: Tennessee Comprehensive Assessment Program (TCAP) **2014**: Field test PARCC **2015**: TNReady	**2014**: Tennessee Comprehensive Assessment Program (TCAP) **2014**: Field test PARCC **2015**: TNReady
Texas	**2015**: State of Texas Assessments of Academic Readiness (STAAR)	**2015**: STARR
Utah	**2014**: Field test Student Assessment of Growth and Excellence (SAGE) **2015**: SAGE	**2014**: Field test Student Assessment of Growth and Excellence (SAGE) **2015**: SAGE, ACT
Vermont	**2010**: New England Common Assessment Program (NECAP) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: New England Common Assessment Program (NECAP) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Virginia	**2010**: Standards of Learning (SOL)	**2010**: SOL

Table A4.18 - Aligned Testing Details (continued)

State	3-8 grades	High school
Washington	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2010**: High School Proficiency Exam **2014**: Field test Smarter Balanced **2015**: Smarter Balanced
West Virginia	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced	**2014**: Field test Smarter Balanced **2015**: Smarter Balanced
Wisconsin	**2010**: Wisconsin Knowledge and Concepts Exam (WKCE) **2014**: Field test Smarter Balanced **2015**: Smarter Balanced **2016**: Wisconsin Forward	**2015**: ACT, ACT Aspire
Wyoming	**2010**: Proficiency Assessments for Wyoming Students (PAWS), Student Assessment of Writing Skills (SAWS) **2013**: Field test PAWS revised **2014**: PAWS revised, Field test Smarter Balanced **2017**: Wyoming Test of Proficiency and Progress (WY-TOPP)	**2016**: ACT Aspire (9-10), ACT (11) **2017**: Wyoming Test of Proficiency and Progress (WY-TOPP), ACT

Table A4.19: Treatment codings by state

State	Main treatment: CCSS Adoption	CCSS implementation requirement	CCSS implementation strategies	Effective CCSS implementation	Include temporary CCSS adopters and implementers	CC-aligned testing
Alabama	2010	2013	2010	Always Control	2010	2014
Alaska	Always Control	Always Control	Always Control	Always Control	Always Control	Always Control
Arizona	2010	2013	2010	2010	2010	2014
Arkansas	2010	2013	2010	2010	2010	2014
California	2010	2014	2010	2010	2010	2014
Colorado	2010	2013	2010	2010	2010	2012
Connecticut	2010	2013	2010	2010	2010	2014
Delaware	2010	2012	2010	2010	2010	2014
D.C.	2010	2012	Always Control	2010	2010	2012
Florida	Always Control	Always Control	2010	Always Control	2010	Always Control
Georgia	2010	2014	2010	Always Control	2010	2015
Hawaii	2010	2013	2010	2010	2010	2014
Idaho	2011	2013	2010	2011	2011	2013
Illinois	2010	2013	2010	2010	2010	2014
Indiana	Always Control	Always Control	2010	Always Control	2010	Always Control
Iowa	2010	2012	2010	2010	2010	2011
Kansas	2010	2013	2010	2010	2010	2015
Kentucky	2010	2011	2010	2010	2010	2010
Louisiana	2010	2013	Always Control	2010	2010	2013
Maine	2011	2012	2010	2011	2011	2010
Maryland	2010	2013	2010	2010	2010	2014
Massachusetts	2010	2013	2010	Always Control	2010	2010
Michigan	2010	2012	2010	2010	2010	2010
Minnesota	2010	2012	Always Control	2010	2010	Always Control
Mississippi	2010	2013	2010	2010	2010	2014
Missouri	2010	2014	2010	2010	2010	2014
Montana	2011	2013	Always Control	2011	2011	2014

Treatment codings by state (continued)

State	Main treatment: CCSS Adoption	CCSS implementation requirement	CCSS implementation strategies	Effective CCSS implementation	Include temporary CCSS adopters and implementers	CC-aligned testing
Nebraska	Always Control	Always Control	Always Control	Always Control	Always Control	Always Control
Nevada	2010	2013	2010	2010	2010	2014
New Hampshire	2010	2014	2010	2010	2010	2013
New Jersey	2010	2013	2010	2010	2010	2014
New Mexico	2010	2013	2010	2010	2010	2014
New York	2010	2013	2010	2010	2010	2012
North Carolina	2010	2012	2010	2010	2010	2012
North Dakota	2011	2013	2010	2011	2010	2015
Ohio	2010	2013	2010	2010	2010	2014
Oklahoma	Always Control	Always Control	2010	Always Control	2010	Always Control
Oregon	2010	2014	2010	2010	2010	2014
Pennsylvania	2010	2013	2010	2010	2010	2013
Rhode Island	2010	2013	2010	2010	2010	2014
South Carolina	Always Control	Always Control	2010	Always Control	2010	Always Control
South Dakota	2010	2014	Always Control	2010	2010	2014
Tennessee	2010	2013	2010	2010	2010	2014
Texas	Always Control	Always Control	Always Control	Always Control	Always Control	Always Control
Utah	2010	2013	2010	2010	2010	2014
Vermont	2010	2013	2010	2010	2010	2014
Virginia	Always Control	Always Control	Always Control	Always Control	Always Control	Always Control
Washington	2011	2014	Always Control	2011	2012	2014
West Virginia	2010	2014	2010	2010	2010	2014
Wisconsin	2010	2014	2010	2010	2010	2014
Wyoming	2012	2014	Always Control	2012	2012	2013

Notes: Table shows whether schooling years in a given state are never coded as exposed to CCSS ("always control"), or, if they are, from which year onwards, for different treatment definitions. Data sources: Achieve Inc. (2013) and CCSSI (2022) (CCSS Adoption and CCSS implementation requirements); Webber et al. (2014) (CCSS implementation strategies); Carmichael et al. (2010) (Effective CCSS implementation); Bleiberg (2021) (Include temporary CCSS adopters and implementers); Own research, see also A4.18 (CCSS-aligned testing)

5 Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments*

5.1 Introduction

Education is a core determinant of life outcomes, both for individuals and societies at large (Krueger and Lindahl, 2001; Hanushek and Woessmann, 2008; Acemoglu and Autor, 2011). Hence, improving equity and efficiency in education systems is a central policy goal in modern societies. To achieve such improvements, it is important to understand the role of genetic endowments for educational attainment: on the one hand, genetic endowments are strong predictors of education; in heritability studies they account for 40% of the variation in years of education (Branigan et al., 2013; Lee et al., 2018). On the other hand, the importance of genetic endowments varies with social environments like families, neighborhoods, and schools (Cesarini and Visscher, 2017; Koellinger and Harden, 2018). Therefore, the link between genetic endowments and life outcomes may be modified by policy interventions. This observation raises important questions: can school reforms moderate the link between genetic endowments and educational outcomes? If yes, which domains of school environments are particularly effective in doing so? Answers to these questions are of utmost importance to address equity and efficiency concerns in the production of educational attainment. Despite this importance, current evidence is scant.

In this paper, we study the interaction of genetic endowments and school environments in the production of educational attainment. We focus on two dimensions of school environments that have been studied extensively in the literature on education economics: teacher quality and class size (Angrist and Lavy, 1999; Rockoff, 2004; Rivkin et al., 2005; Fredriksson et al., 2013; Chetty et al., 2014a,b; Jackson, 2018; Angrist et al., 2019; Leuven and Løkken, 2020). Furthermore, these dimensions can be directly influenced by policymakers but their reform applies to all children and does not presuppose any form of genetic screening—a practice many of us would be uncomfortable with (Martschenko et al., 2019).

We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to study the interaction of genetic endowments and school environments in a between-family design. Add Health is a 5-wave panel study that follows a representative sample of US high school students from 1994/95 until the present day. To the best of our knowledge, Add Health is the only (publicly available) data set that offers detailed information on schooling environ-

* This chapter is joint work with Paul Hufe and Marc Stoeckli. It is based on the paper 'Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments', mimeo.

ments from both survey and administrative sources for a genotyped sample of reasonable size.

To measure genetic endowments, we leverage recent advances in molecular biology and use a polygenic score for educational attainment (PGS^{EA}, Dudbridge, 2013; Lee et al., 2018). PGS^{EA} is an individual measure for the genetic propensity to attain education.[1] It is fixed at conception and cannot be modified by environmental interventions thereafter. Therefore, PGS^{EA} confers important advantages over traditional proxies for "innate ability", such as student test scores and IQ tests (Hanushek and Woessmann, 2008; Heckman et al., 2010; Brinch and Galloway, 2012; Hanushek and Woessmann, 2012). To measure the quality of school environments, we use information from headmaster surveys and administrative data sources such as the Common Core of Data, and conduct a principal component analysis on the following school-level characteristics: teacher experience, teacher turnover, teacher education, teacher diversity as well as class sizes and student-teacher ratios. From this analysis, we extract two factors that are indicative for the quality of teachers ($I_{Quality}$) and the quantity of teachers relative to the number of students ($I_{Quantity}$), respectively.

Clean causal identification of gene-environment interactions is challenging. In this study, we rely on a between-family comparison in which we control for an extensive set of predetermined family background characteristics. We discuss the associated identification assumptions in detail and provide tests for their satisfaction. First, while genetic endowments are fixed at conception they are correlated with other family characteristics that co-determine educational attainment. Therefore, our parameters of interest may be confounded by *genetic nurture effects*. In response, we show that the relevant point estimates from the between-family design replicate in a smaller sibling sample that allows us to control for genetic nurture by including family fixed effects. Second, school characteristics may be correlated with other family characteristics that co-determine educational attainment. Therefore, our parameters of interest may be confounded by *selection effects*. In response, we show that our main findings are robust to bounding exercises à la Oster (2019) and Cinelli and Hazlett (2020). Lastly, gene-environment interactions can only be identified if genetic endowments and the environmental variable of interest are distributed independently of each other. In response, we show that we cannot reject the equality of PGS^{EA} distributions in various school environments. In summary, although the between-family design does not allow to cleanly identify causal effects, all our empirical tests point towards the satisfaction of the relevant identification requirements. In addition, our results withstand a series of empirical tests for competing mechanisms that we discuss in detail below.

Our results can be summarized as follows. First, genetic endowments and teacher quality are highly predictive for years of education: a one-standard-deviation increase in PGS^{EA} (teacher

[1] In addition, PGS^{EA} has been shown to be highly predictive for a number of life outcomes that are closely related to educational attainment. These outcomes include earnings, wealth and (non-)cognitive skills (Lee et al., 2018; Barth et al., 2020; Houmark et al., 2020; Muslimova et al., 2020; Papageorge and Thom, 2020; Buser et al., 2021a; Demange et al., 2021).

quality) increases educational attainment by ≈ 0.37 (0.22) years. These increases can be compared to a sample average of 14.81 years and correspond to 16.44% (9.8%) of a standard deviation. Second, genetic endowments and teacher quality act as substitutes in the production of educational attainment: a one-standard-deviation increase in teacher quality reduces the positive association of educational attainment with PGS^{EA} by ≈ 19%. This result implies that improvements in the quality of teachers may reduce the genetic gradient in educational attainment. Furthermore, it suggests that teacher quality may countervail the effects of family socio-economic status—an environmental characteristic that tends to magnify the genetic gradient in educational attainment (Papageorge and Thom, 2020; Ronda et al., forthcoming).[2] Third, in contrast to teacher quality, teacher quantity is not associated with educational attainment—a null result that does not vary across the PGS^{EA} distribution.

We perform a series of robustness checks to evaluate whether our results are conflated by competing mechanisms. We begin by showing that our measures for teacher quality and teacher quantity do not pick up the effects of other school characteristics that may correlate with student outcomes. These characteristics comprise school peer characteristics, school-level policies such as sanctions for academic misconduct, and overall school value-added. Next, we demonstrate that our results are not driven by gene-environment interactions that reflect family instead of school environments. To that end, we run a fully interacted model controlling for all possible interactions between PGS^{EA}, $I_{Quality}$, $I_{Quantity}$, and a broad set of parental background characteristics (Keller, 2014). In addition, we show that there is no differential association between PGS^{EA} and parental time investments depending on school quality.

We also analyze the mechanisms that underpin the substitutability of genetic endowments and teacher quality. Educational attainment summarizes information from various educational stages, where each stage requires a different mix of skills (Cunha et al., 2006, 2010). Therefore, we repeat our analysis by replacing total educational attainment with binary variables for whether or not respondents achieved a given educational degree. We find that the substitutability of genetic endowments and teacher quality is largest at the stage of college education. In contrast, we find little substitutability for the probability of graduating from high school, and none for the probability of obtaining post-graduate degrees. These results provide a notable contrast to Papageorge and Thom (2020), who find a growing complementarity of parental background characteristics and genetic endowments as individuals progress through the educational system. To uncover which type of skills drives our results, we analyze the associations of PGS^{EA} and teacher quality with a set of intermediate outcomes including subjective and objective health, cognitive skills, economic preferences, and personality measures. We find that substitutabilities of genetic endowments and teacher quality with respect to subjective health, verbal intelligence, risk-aversion, and patience underpin our main result.

[2] See also our replication of their findings in section 5.5.

Our study contributes to three strands of literature. First, we contribute to the literature on gene-environment interactions. Existing evidence shows that the association between socio-economic outcomes and genetic endowments varies with the socio-economic status of parents (Houmark et al., 2020; Papageorge and Thom, 2020; Ronda et al., forthcoming). Evidence on gene-environment interactions regarding school environments is more scant. Barcellos et al. (2021) use a compulsory schooling reform to show that returns to schooling are lower for genetically advantaged students. However, they focus on the length of education and not the quality of school environments. Trejo et al. (2018) show a stronger genetic gradient in schools with better educated parents. However, in the presence of endogenous sorting the composition of schools is difficult to control. Therefore, we focus on margins that can be directly targeted by policymakers: the quality and quantity of teachers.

Second, we contribute to the literature on teacher quality. The positive effects of teacher quality on short- and long-term outcomes of students are well-documented (Rockoff, 2004; Rivkin et al., 2005; Chetty et al., 2014a,b; Jackson, 2018). However, the literature is far less conclusive about the equalizing effect of teacher quality across student subgroups. For example, Aaronson et al. (2007) find that low achieving students benefit more from high-quality teachers. In contrast, Chetty et al. (2014b) show that students from minority and low-income backgrounds benefit less. While existing studies have evaluated heterogeneities along dimensions that conflate genetic and social factors, we are able to measure the genetic predisposition for educational success as fixed at conception. We show that investments in the quality of teachers cushion the genetic gradient in educational attainment.

Third, we contribute to the literature on class size. Here, the average effects on students outcomes are subject to academic debate. On the one hand, experimental studies on class size reductions tend to find positive effects on student achievement (Krueger, 1999; Chetty et al., 2011). On the other hand, quasi-experimental analyses exploiting maximum class-size rules tend to find mixed results even if they analyse similar settings (Angrist and Lavy, 1999; Fredriksson et al., 2013; Angrist et al., 2019; Leuven and Løkken, 2020). Similarly, the equalizing effects of class-size reductions are contested. For example, Krueger (1999) shows that class size reductions are more beneficial to students from minority and low-income background. In contrast, Fredriksson et al. (2013) document that wage increases following class size reductions are more pronounced for students from high income backgrounds. Our study is the first to evaluate heterogeneities along the genetic dimension. We show that teacher quantity is not associated with gains in educational attainment irrespective of genetic endowment.

Our results are policy relevant. First, we show that higher-quality teachers are conducive to educational attainment in the lower tail of the PGS^{EA} distribution but do not compromise achievement in the upper tail. This finding suggests that policymakers do not face an equity-efficiency trade-off when investing into the quality of teachers. Second, in contrast to teacher quality, we find no effect of teacher quantity on the educational outcomes of students irrespective of their genetic endowments. This finding suggests that policymakers who are willing

to address the equity and efficiency concerns related to genetic endowments do not face a trade-off between investments into teacher quality and teacher quantity. This last finding is economically relevant as salaries and employee benefits of teachers are by far the largest cost factor in the US school system, accounting for about half of the expenditures in US public primary and secondary schools (Figure A5.1).

The remainder of this paper is structured as follows. In section 5.2, we provide a primer on the measurement of genetic endowments. In section 5.3, we detail our empirical strategy. After introducing our data sources in section 5.4, we present results in section 5.5. Section 5.6 concludes the paper.

5.2 Measuring Genetic Endowment

The "First Law of Behavior Genetics" states that all human behavioral traits are heritable (Turkheimer, 2000). That is, genetic endowments explain the expression of any trait at least to some extent. The empirical challenge is to identify the specific sequences in the genome that are related to the traits of interest.[3] Recent advances in molecular genetics have enabled a novel method of genetic discovery: genome-wide association studies (GWAS). GWAS exploit the most common type of genetic variation between humans, so-called single-nucleotide polymorphisms (SNP). SNPs occur when a single nucleotide—the basic building block of DNA molecules—differs at a specific position in the genome. Humans have between four and five million SNPs. GWAS estimate separate linear regressions that relate a SNP of individual i at genome location j to an outcome of interest y:

$$y_i = \psi_j^y SNP_{ij} + \delta C_i + \varepsilon_i. \tag{5.1}$$

$SNP_{ij} \in \{0, 1, 2\}$ is a count variable and indicates the number of minor allele that individual i possesses at location j. Minor alleles are the less frequent genetic variation within a population. As humans inherit one of each chromosome from each parent, they possess either zero, one, or two minor alleles at each location j. C_i is a vector of control variables to filter out spurious correlations due to non-biological differences across population groups. A particular SNP coefficient ψ_j^y is referred to as genome-wide significant if the null hypothesis of non-association is rejected at a level of $p < 5 \times 10^{-8}$ (Chanock et al., 2007). The p-value is deliberately low to adjust for multiple hypothesis testing.

The association of any single SNP with y is minuscule, but jointly they can explain a substantial share of the observed outcome differences between individuals (Lee et al., 2018). In particular, the estimated SNP coefficients can be used to construct polygenic scores (PGS). A PGS is a single quantitative measure of an individual's genetic propensity toward an outcome relative

[3] Human genetic information is stored in 23 chromosome pairs that consist of deoxyribonucleic acid (DNA) molecules. These chromosomes, in turn, contain 20,000 to 25,000 genes—specific DNA sequences that provide instructions for building particular proteins. About 99% of the sequences are identical across humans.

to the population. Formally, individual i's PGS for outcome y, PGS_i^y, is constructed by linearly aggregating all SNP_{ij} using ψ_j^y as weighting factors:

$$PGS_i^y = \sum_j \widehat{\psi}_j^y SNP_{ij}, \tag{5.2}$$

where $\widehat{\psi}_j^y$ is the estimated SNP coefficient from equation (5.1). To avoid overfitting, equation (5.1) is estimated in a discovery sample, whereas the PGS is constructed in a hold-out sample (Wray et al., 2014).

The predictive power of a PGS is broadly determined by two factors: the heritability of the outcome, which serves as an upper bound of the variance the PGS can explain; and the size of the discovery sample (Dudbridge, 2013). All else equal, the more heritable the outcome, or the larger the discovery sample to estimate the aggregation weights $\widehat{\psi}_j^y$, the higher the predictive accuracy of the PGS. For example, the heritability of educational attainment is around 40% (Branigan et al., 2013). The PGS for educational attainment constructed by Lee et al. (2018) is based on information from 1.1 million individuals and explains 12.7% of the variance in educational attainment.

The interpretation of PGS is non-trivial. First, PGS are not purely measures of biological influence. In particular, GWAS coefficients may capture environmental factors such as population stratification across geographic regions (Abdellaoui et al., 2019). To this address this concern, we follow standard practice and always control for the first 20 principal components of the genetic data in our empirical analysis.[4] Second, the explanatory power of PGS is contingent on the context of its application. If a PGS is applied in one context, whereas the underlying GWAS was estimated in a very different context, the predictive power of the PGS will be attenuated. In our context, this concern is limited: we apply PGS to a sample from the United States, whereas the underlying GWAS predominantly draw on samples from other industrialized countries with comparable education systems. Third, PGS are noisy measures of genetic endowments. Due to current GWAS sample sizes, they do not capture all genetic variation relevant for the outcome of interest. As a direct consequence, alternative PGS are still predictive for educational attainment over and above PGS^{EA}. However, in Appendix Table A5.4, we show that PGS^{EA} is significantly more predictive than any plausible alternative PGS. Therefore, it is the best among other noisy measures for genetic endowments.

PGS are now available for a wide variety of outcomes, including body mass index and height (Yengo et al., 2018), attention deficit hyperactivity disorder (Demontis et al., 2019), major depressive disorder (Howard et al., 2019), intelligence (Savage et al., 2018), smoking (Liu et al., 2019), and sleep duration (Jansen et al., 2019). For our analysis, we rely on the PGS for educational attainment by Lee et al. (2018).

[4] The first principal components of the full matrix of genetic data capture most of the geographical variation in allele frequencies (see Mills et al., 2020, chapter 9.4, for a discussion). Therefore, they control for the geographic correlation between allele frequencies and socio-economic status.

5.3 Empirical Strategy

5.3.1 Empirical Model

Consider a model in which skills θ of child i at age a are determined by prior skill levels θ_{ia-1}, parental investments I^P_{ia}, school investments I^S_{ia}, and genetic endowments G_i.[5] There are three phases of skill accumulation:

$$\theta_{ia} = \begin{cases} f_a(G_i) & \text{, for child age } a = 0, \\ f_a(\theta_{ia-1}, I^P_{ia}, G_i) & \text{, for child age } a = 1, ..., 5, \\ f_a(\theta_{ia-1}, I^P_{ia}, I^S_{ia}, G_i) & \text{, for child age } a = 6, ..., A. \end{cases} \tag{5.3}$$

That is, skills at conception are determined by genetic endowments only. For child ages $a = 1, ..., 5$, i.e. prior to attending school, parents are the only source of investments into skills in this model. Parental investments include monetary investments such as buying toys or books but also time investments such as playing with or talking to the child. For $a = 6, ..., A$, schools are an additional source of investments into skills. School-based investments include instruction by teachers or interactions with peers.

Furthermore, assume completed education Y to be a function of individual skills accumulated by the end of childhood at age $a = A$:

$$Y_i = g(\theta_{iA}). \tag{5.4}$$

Recursively substituting equations (5.3) and (5.4) across child ages $a = 1, ..., A$, we obtain a model in which educational attainment is determined by initial genetic endowments, the history of family inputs, and the history of schooling inputs:

$$Y_i = h(I^P_{iA}, ..., I^P_{i1}, I^S_{iA}, ..., I^S_{i6}, G_i). \tag{5.5}$$

We are interested in the complementarity of schooling inputs and genetic endowments at a particular child age a:

$$\kappa = \frac{\partial^2 h(\overline{I^P_{ia}}, \overline{I^P_{ia-1}}, ..., \overline{I^P_{i1}}, I^S_{ia}, \overline{I^S_{ia-1}}, ..., \overline{I^S_{i6}}, G_i)}{\partial I^S_{ia} \partial G_i}. \tag{5.6}$$

[5] For the sake of simplicity, we abstract from other actors in the child development process.

If $\kappa < 0$, genetic endowments and school investments at age a are *substitutes* in the production of educational attainment. School investments are then less productive for individuals with high genetic endowments. Reversely, if $\kappa > 0$, genetic endowments and school investments at age a are *complements* in the production of educational attainment. School investments are then more productive for individuals with high genetic endowments.

In this study, we focus on school investments during high school ($14 \leq a \leq 18$). We estimate the complementarity parameter κ from a linear regression model with an interaction term:

$$Y_i = \alpha G_i + \beta I_{ia}^S + \kappa(G_i \times I_{ia}^S) + \mathbf{X}_i(a)\gamma + \epsilon_i, \tag{5.7}$$

where $\mathbf{X}_i(a)$ denotes a vector of control variables to condition on the history of family and schooling inputs until age $a = 14$.

5.3.2 Identification

Unbiased estimation of κ is based on the following set of requirements: (i) the effect of G_i is identified, (ii) the effect of I_{ia}^S is identified, and (iii) G_i and I_{ia}^S are assigned independently from each other (Almond and Mazumder, 2013; Nicoletti and Rabe, 2014; Johnson and Jackson, 2019). In the following, we will discuss each of these requirements, potential threats to their satisfaction, and how we address them in the context of this paper.

(i) Absence of genetic nurture effects. Genetic endowments are fixed at conception, yet they are not exogenous to family characteristics that co-determine educational attainment. During meiosis, genetic endowments of children are randomly drawn from the genetic pool of their biological parents.[6] As a consequence, G_i is a function of maternal and paternal genetic endowments. These parental endowments, however, may also correlate with parental investments $I_{i1}^P, ..., I_{ia}^P$. Hence, in estimating equation (5.7), α and κ may be confounded by *genetic nurture effects* (Kong et al., 2018). Genetic nurture can be controlled either by estimating a sibling fixed effects model that relies on within-family variation in G_i only (Selzam et al., 2019; Houmark et al., 2020; Kweon et al., 2020); in a non-transmitted genes design, where one includes both maternal and paternal genetic endowments in control vector $\mathbf{X}_i(a)$; or in an adoption design, where offspring are biologically unrelated to their parents (see Demange et al., 2020, for a detailed comparison of all three approaches). All approaches, however, are very data demanding. For example, the sibling design requires the availability of both a large set of siblings and individual measurements of G_i. Therefore, it can only be applied in a very limited set of existing data sets.

In this study, we estimate a between-family model in which we use an extensive set of pre-determined family background characteristics to control for genetic nurture effects. This

[6] In this process, chromosomes of fathers and mothers are re-combined to produce genetically distinct offspring. Therefore, singleton children of the same parents are never genetically identical to their siblings. Furthermore, conditional on the parents' genome, the offspring's set of genes is randomly distributed.

approach is standard in the literature and intends to approximate requirement (i) while maximizing statistical power to detect the sought-after gene-environment interaction (Domingue et al., 2020). Reassuringly, controlling for $\mathbf{X}_i(a)$ in our between-family model, we obtain a point estimate of α that replicates the corresponding estimate from a sibling fixed effects model on a subsample of our data (N = 525).

(ii) Absence of selection effects. Parents choose schools based on school characteristics. Therefore, the latter may not be exogenous to family characteristics that co-determine educational attainment (Altonji et al., 2005; Beuermann et al., 2018). As a consequence, I_{ia}^S is a function of observed and unobserved family and child characteristics that may correlate with parental investments $I_{i1}^P, ..., I_{ia}^P$. Hence, in estimation model (5.7), β and κ may be confounded by *selection effects* (Altonji et al., 2005; Altonji and Mansfield, 2018; Biasi, forthcoming). Selection into schools can be controlled in (quasi-)experimental settings, e.g. using variation based on admission lotteries (Cullen et al., 2006; Angrist et al., 2016), or the geographic design of catchment areas (Laliberte, 2021). Existing data sets that avail such variation, however, do not contain sequenced DNA data that are required to measure G_i at the individual level.

In this study, we use an extensive set of pre-determined family background characteristics to control for selection into schools based on observables. To assess potential confounding through selection based on unobservables we calculate *bias-adjusted treatment effects* along the lines of Oster (2019) and Cinelli and Hazlett (2020). Both procedures assume that changes in the coefficients of interest due to the introduction of observables are informative for the extent of confounding due to unobservables. Reassuringly, applying these correction methods our results for β and κ remain qualitatively unaltered in comparison to our benchmark estimates.

(iii) Independent assignment of genetic endowments and school environments. Requirements (i) and (ii) must be combined such that G_i and I_{ia}^S are distributed independently of each other. Strong correlation between G_i and I_{ia}^S implies little overlap in the distributions of G_i at different levels of I_{ia}^S and vice versa. As a consequence, two empirical challenges arise. First, there may not be sufficient variation to identify $\alpha, \beta, and \kappa$ separately from each other. Second, κ would be identified from the tails in the respective distributions. One then would always compare individuals with similar G_i that score unusually high or low in their school-quality specific distribution of genetic endowments and vice versa. Among others, these concerns would be addressed in a setting that avails (quasi-)experimental variation in I_{ia}^S at the level of siblings from the same biological parents. However, as highlighted in our previous discussion, such a setting is unlikely to be found in existing data sources.

To verify the satisfaction of requirement (iii), we present empirical evidence that G_i and I_{ia}^S are indeed distributed independently of each other. This conclusion holds both unconditionally and controlling for $\mathbf{X}_i(a)$.

In summary: in an ideal setting, one would estimate the complementarity parameter κ by combining a sibling fixed effects model with experimental variation in school characteristics

among children of the same biological parents. To date, there is no single data set that simultaneously avails genetic data at the individual level, a large set of siblings, and quasi-experimental variation in school assignment. Therefore, we approximate the conditions of such an ideal setting with the best data available to us. Our estimates of $\alpha, \beta,$ and κ do not have a strict causal interpretation. However, we demonstrate their robustness to a large battery of potential confounders including school peer effects, school sanction policies, parental time investments, and potential non-linearities of genetic effects by other individual characteristics. Furthermore, we show that our baseline estimates of genetic effects are consistent with the estimates from a sibling fixed effects model. Erring on the side of caution, we nevertheless speak of *associations* instead *causal effects* in the remainder of the paper.

5.4 Data

We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a 5-wave panel study that focuses on the determinants of health-related behaviors and health outcomes. Add Health is a nationally representative sample of adolescents enrolled in grades 7–12 in 1994/95. Initial information (wave 1, $N = 20,745$) was collected from a stratified sample of 80 high schools across the US as well as their associated feeder schools. In addition to in-depth interviews with adolescents, questionnaires were administered to school representatives, parents, and roughly 90,000 students of the sampled schools. Follow-up in-home questionnaires were collected in 1996 (wave 2, $N = 14,738$), 2001/02 (wave 3, $N = 15,179$), and 2008/09 (wave 4, $N = 15,701$). In the most recent wave (wave 5, 2016/18, $N = 12,300$), Add Health respondents are between 33 and 43 years old.

In the following, we describe our main variables of interest. Detailed descriptions of all variables used in our analysis are disclosed in Supplementary Material A5.3.

Outcomes. We measure educational attainment Y_i by total years of education. In each wave, respondents were asked about their highest level of education at the time of the interview. For each individual, we use the most recent information and transform education levels into years of education following the mapping suggested by Domingue et al. (2015).[7]

To analyze the mechanisms behind our headline results, we additionally use a series of measures for academic degrees, health, and (non-)cognitive skills. First, academic degrees allow us to investigate at which educational stage our results emerge. We focus on whether respondents finished high school, obtained a college degree, or obtained a post-graduate degree.

[7] Numeric values in parentheses: eighth grade or less (8), some high school (10), high school graduate (12), GED (12), some vocational/technical training (13), some community college (14), some college (14), completed vocational/technical training (14), associate or junior college degree (14), completed college (16), some graduate school (17), completed a master's degree (18), some post-baccalaureate professional education (18), some graduate training beyond a master's degree (19), completed post-baccalaureate professional education (19), completed a doctoral degree (20).

Second, measures for health and (non-)cognitive skills serve as proxy variables of θ_{iA} and allow us to analyze the dimensions of skill development that drive the main findings on educational attainment. We proxy health by quality-adjusted life years (QALY) that we derive from self-assessed health measures as well as a summary index of diagnosed health conditions. We proxy cognitive skills by the Picture Vocabulary Test (PVT), a test for receptive hearing vocabulary that is a widely-used proxy for verbal ability and scholastic aptitude. We proxy non-cognitive skills by self-reported measures of general risk aversion and patience (Falk et al., 2018) as well as self-reported information on the Big Five personality traits (Almlund et al., 2011).

Genetic endowments. Add Health obtained saliva samples from consenting participants in wave 4. After quality control procedures, genotyped data is available for 9,974 individuals and 609,130 SNPs. Add Health uses this data to calculate different PGS using summary statistics from existing GWAS. We use a PGS for educational attainment, denoted by PGS^{EA}, that is based on the GWAS of Lee et al. (2018).[8]

Lee et al. (2018) perform a meta-analysis of 71 quality-controlled cohort-level GWAS. Their meta-analysis produced association statistics for around 10 million SNPs, of which 1,271 reached genome-wide significance. Genes near these genome-wide significant SNPs are relevant for the central nervous system, and many of them encode proteins that carry out neurophysiological functions such as neurotransmitter secretion or synaptic plasticity. They are relevant for brain-development processes prior to and after birth.

PGS^{EA} is highly predictive for educational attainment and has been widely used in existing studies. Lee et al. (2018) suggest that PGS^{EA} is a better predictor for years of education than household income. Including the score in a regression of years of education on a set of controls yields an incremental R^2 of 0.127 in the Add Health sample. Among others, PGS^{EA} has been used to study the formation of early childhood skills (Belsky et al., 2016), educational attainment (Domingue et al., 2015; Houmark et al., 2020), earnings (Papageorge and Thom, 2020), wealth accumulation (Barth et al., 2020), and social mobility (Belsky et al., 2018).

We standardize PGS^{EA} on our analysis sample to have a mean of zero ($\mu = 0$) and a standard deviation of one ($\sigma = 1$).

School investments. In wave 1 and 2, Add Health administered detailed questionnaires to headmasters of Add Health schools. The schools are also linked to administrative data from the Common Core of Data (CCD) and the Private School Survey (PSS). We use these sources to construct indicators for I_{ia}^{S} using a principal component analysis that includes the following school-level information: (i) average class size, (ii) average student-teacher ratio, (iii) share

[8] Lee et al. (2018) construct PGS^{EA} for two prediction cohorts, Add Health and the Health and Retirement Study (HRS). PGS^{EA} is based on results from the meta-analysis that excluded these two cohorts from the discovery sample. PGS^{EA} was generated from HapMap3 SNPs using the software LDpred, a Bayesian method that weights each SNP by the posterior mean of its conditional effect, given other SNPs.

of teachers with a master degree, (iv) share of new teachers in the current school year, (v) share of teachers with school-specific tenure of more than five years, and Herfindahl indices measuring teacher diversity with respect to (vi) race and (vii) Hispanic background.

Many of these characteristics have been shown to predict teacher value-added. For example, Ronfeldt et al. (2013) and Hanushek et al. (2016) show that a high teacher turnover, which we proxy by the share of new teachers, harms the quality of instruction and student achievement. Rockoff (2004) and Papay and Kraft (2015) show that teaching experience, which we proxy by the share of teachers with more than five years of tenure, correlates with teacher performance.[9] Finally, Clotfelter et al. (2010) and Jacob et al. (2018) show that academic credentials, which we proxy by the share of teachers with a master degree, are positively associated with teacher effectiveness.

Figure 5.1 shows the rotated loadings on the first two principal components. The first component loads almost exclusively on average class size and average student-teacher ratio. Hence, we interpret this component as an indicator for the "quantity" of teachers, denoted by $I_{Quantity}$. The second component loads positively on the percentage of teachers with a master degree and the share of teachers with a tenure of more than five years; it loads negatively on the share of new teachers in the current school year. We interpret this component as an indicator for the "quality" of teachers, denoted by $I_{Quality}$. Both factors are coded such that higher values indicate higher school investments, i.e. higher teacher "quantity" investments (smaller classes) and higher teacher "quality" investments (better teachers), respectively. The calculated factors are orthogonal to each other by construction and standardized to $\mu = 0$ and $\sigma = 1$.[10]

Control variables. Add Health provides extensive information about the environments that respondents were exposed to during childhood. We approximate the identification pre-requisites discussed in section 5.3 by choosing a vector of pre-determined variables $\mathbf{X}_i(a)$ to control for genetic nurture effects and selection into schools. Specifically, we control for family background characteristics by including maternal and paternal education (in years), the family's religious affiliation (Christian/Non-Christian), and maternal age at birth (in years). Furthermore, we include the mean and standard deviation of potential wages for both mother and father across child ages 0–14.[11] At the level of children, we control for age in months,

[9] These teachers have taught for at least five years in their life and hence do not suffer from a lack of basic teaching experience. Since we measure tenure at the current school, the measure combines information about teaching experience with information about teacher turnover.

[10] Intuitively, one may expect a negative correlation between teacher quality and quantity: conditional on a given budget, a school administrator may prefer to invest in teacher quality at the expense of average class sizes or vice versa. However, this is not what we observe in the data. If quality and quantity were substitutes, we would expect loadings on the two principal components to pull into diametrically opposed directions. To the contrary, we find that the variables capturing the quality and quantity dimensions are orthogonal to each other and almost exclusively load on one principal component only.

[11] Note that Add Health contains information on actual income. However, actual income may be a bad control as it reflects parental responses to both G_i and I_{ia}^S. Therefore, we follow the procedure of Shenhav (2021) and combine the 1970 Census and the March Current Population Survey (1975–2000) to construct potential wages

sex as well as their interaction. We follow standard practice in the literature and account for population stratification in genetic endowments by including the first 20 principal components of the full matrix of genetic data. Lastly, all estimations include a vector of state fixed effects.

Note that we focus on pre-determined variables—variables that are fixed prior to the period of observation—to avoid smearing through "bad controls" (Angrist and Pischke, 2009). However, in robustness analyses we expand the vector of controls by potentially endogenous parental time investments and family income. Our results remain unaffected.

Analysis sample. We apply the following sample selection criteria. First, we restrict our sample to genotyped respondents of European descent.[12] This is common practice in the literature because GWAS are predominantly conducted on this ancestry group. Conducting GWAS on just one ancestry group yields low statistical power to account for population stratification between groups, leading to biased estimates of genetic influence (Martin et al., 2017; Ware et al., 2017).[13]

Second, we retain the subsample of individuals who (i) visited an Add Health high school or an associated feeder school in wave 1, and (ii) for whom the high school exit record indicates that they had graduated from the same school. These sample selection criteria strike a balance between sample size and the matching accuracy of individuals with our measures for schooling environments. For example, imposing criterion (i) we assume that individuals indeed transition from feeder schools to designated Add Health schools. Thereby, we increase our sample but may erroneously assign information on I_{ia}^S to individuals transitioning to high schools out of the Add Health universe. Reversely, imposing criterion (ii) we exclude individuals that may have moved to other high schools throughout grades 9–12. Thereby, we reduce our sample size but minimize the risk of erroneously assigning information on I_{ia}^S to movers. We note that neither strengthening (i) by excluding individuals from feeder schools, nor relaxing (ii) by assuming that individuals remain at the same school across grades 9–12 overturns our main conclusions (Appendix Table A5.5).

Third, we drop all observations with missing information in Y_i, G_i, I_{ia}^S, and $\mathbf{X}_i(a)$ through list-wise deletion.

Applying these restrictions, we obtain a sample of $3,081$ individuals from 77 high schools across the US for which we provide summary statistics in Table 5.1. 55% are female, and the average age measured at wave 1 equals ≈ 16 years (194 months). The average educational attainment in our sample is 14.8 years which exceeds the average educational attainment in the parental generation by ≈ 1.1 years. Almost all individuals graduate from high school,

for gender/education/census region/race/ethnicity cells and match these potential wages to parents at each child age $a = 1, ..., 14$.

[12] Ancestry groups in Add Health are identified by principal component analysis on all unrelated members of the full Add Health genotyped sample.

[13] See also the discussion on genetic nurture effects in section 5.3.

which is not surprising given our sample restriction to individuals of European descent who stayed at the same high school in grades 9–12. The college completion rate equals $\approx 50\%$.

To assess the sample representativity, we compare our analysis sample to the 1974–1983 birth cohorts of Non-Hispanic Whites in the American Community Survey (ACS) and the Current Population Survey (CPS) (Appendix Table A5.1). This comparison shows a slight overrepresentation of females and children from young mothers in our sample. Otherwise, our sample is by-and-large comparable to the corresponding groups in ACS and CPS. In robustness analyses we re-weight our analysis sample to match ACS and CPS with respect to gender composition, educational attainment of parents, and the age of mothers at birth. Our results remain unaffected (Appendix Table A5.5).

5.5 Results

We present our results in four steps. In section 5.5.1, we discuss the association of educational attainment, genetic endowments, and school investments in light of the identification requirements discussed in section 5.3. In section 5.5.2, we present our estimates for the complementarity parameter κ. After a robustness analysis in section 5.5.3, we conclude with an analysis of mechanisms in section 5.5.4.

5.5.1 The Association of Educational Attainment with Genetic Endowments and School Investments

Figure 5.2 visualizes the association of educational attainment with our measures for genetic endowments G_i and school investments I_{ia}^S. In the left column we show raw correlations that do not account for the control variables $\mathbf{X}_i(a)$. In the right column, we show associations conditional on $\mathbf{X}_i(a)$.

First, PGS^{EA} is highly predictive of educational attainment. Without controls, a one-standard-deviation (1 SD) increase in PGS^{EA} is associated with an increase in educational attainment of 0.608 years. This association does not have a causal interpretation as it may be confounded by genetic nurture effects. When we control for pre-determined child and family characteristics, a 1 SD increase in PGS^{EA} is associated with an increase in educational attainment of 0.380 years. Sibling studies show that genetic nurture effects usually account for 40–50% of the raw association between PGS^{EA} and educational attainment (Selzam et al., 2019; Kweon et al., 2020; Muslimova et al., 2020; Ronda et al., forthcoming). In our case, the association decreases by 38% when we control for child and family background characteristics. This result suggests that $\mathbf{X}_i(a)$ is indeed able to account for genetic nurture effects as confounding factors. This conclusion is further bolstered by a comparison of our between-family model with a sibling fixed effects model that we estimate on a subsample of our data ($N = 525$). In the within-family comparison, which allows us to perfectly control for genetic nurture effects, we obtain

a point estimate of 0.458 that is significant at the 1%-level (Appendix Table A5.2). This point estimate is very close to the result of the between-family comparison controlling for $\mathbf{X}_i(a)$, and lends further credence to our research design.

Second, $I_{Quality}$ is highly predictive of educational attainment. Without controls, a 1 SD increase in $I_{Quality}$ is associated with an increase in educational attainment of 0.541 years. This association does not have a causal interpretation as it may be confounded by selection effects. When we control for pre-determined child and family characteristics, a 1 SD increase in $I_{Quality}$ is associated with an increase in educational attainment of 0.256 years. This 53% decrease reflects positive selection into schools based on "teacher quality"—a pattern that has been thoroughly documented in existing literature for the US (Biasi, forthcoming). Nevertheless, even when accounting for selection, the association of $I_{Quality}$ and educational attainment remains strong and positive. This result confirms prior literature, which has repeatedly demonstrated positive effects of teacher quality on students' educational success (Hanushek and Rivkin, 2010; Chetty et al., 2014a).

Third, $I_{Quantity}$ is not significantly associated with educational attainment. The weakly positive correlation is imprecisely estimated and does not attain statistical significance at conventional levels. Furthermore, this result does not change when accounting for selection effects through the introduction of control vector $\mathbf{X}_i(a)$. This finding is in line with prior literature, which has not been able to establish consistent effects of teacher quantity on students' educational success (Fredriksson et al., 2013; Angrist et al., 2019; Leuven and Løkken, 2020). However, this average association may mask heterogeneity across students with different genetic endowments—a hypothesis that we will test in the following subsection.

Next to genetic nurture effects and selection effects, a high correlation between G_i and I_{ia}^S would pose another threat to identification of the gene-environment interaction. Figure 5.3 suggests that this threat is not operational in our setting. In this figure, we plot the unconditional PGS^{EA} distribution by tercile of $I_{Quality}$ and $I_{Quantity}$, respectively. Visual inspection suggests that PGS^{EA} distributions are almost congruent to each other within each tercile of the two indicators. In Appendix Table A5.3 we present formal statistical tests for this observation. In particular, we residualize PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ using control vector $\mathbf{X}_i(a)$. We then perform two-sample Kolmogorov-Smirnov tests for the equality of PGS^{EA} distributions within the terciles of $I_{Quality}$ and $I_{Quantity}$, respectively. We do not reject the null hypothesis of equal distributions for any of the comparisons at conventional levels of statistical significance. Hence, we conclude that PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ are indeed assigned independently of each other.

5.5.2 The Interplay of Genetic Endowments and School Investments in the Production of Educational Attainment

Table 5.2 shows our baseline estimates for the interaction of genetic endowments and school investments. In all regressions, we include the vector $\mathbf{X}_i(a)$ to control for genetic nurture and selection into schools.

In column (1), we focus on the teacher quality indicator $I_{Quality}$. The point estimates for PGS^{EA} and $I_{Quality}$ replicate the findings from Figure 5.2 and indicate a strong and positive association of PGS^{EA} and $I_{Quality}$ with educational attainment.[14] A 1 SD increase in PGS^{EA} ($I_{Quality}$) increases educational attainment by ≈ 0.37 (≈ 0.23) years.

$PGS^{EA} \times I_{Quality}$ is our estimate for the complementarity parameter κ. The negative coefficient of the interaction term indicates that genetic endowments and teacher quality act as *substitutes* in the production of educational attainment. A 1 SD increase in teacher quality reduces the positive association of educational attainment with PGS^{EA} by ≈ 19% (= 0.07/0.37). This result provides a notable contrast to existing literature investigating the gene-environment interaction between PGS^{EA} and parental socio-economic status which tend to act as *complements* in the production of educational attainment (Papageorge and Thom, 2020; Ronda et al., forthcoming).

In column (2), we focus on $I_{Quantity}$. The point estimate for $I_{Quantity}$ is again statistically indistinguishable from zero. The estimate for $PGS^{EA} \times I_{Quantity}$ indicates that this null result is not driven by heterogeneity along the PGS^{EA} distribution. Our estimate for the complementarity parameter κ is small and not statistically different from zero.

In column (3), we estimate both complementarity parameters in the same model and show that our results remain virtually unchanged. This stability is expected since $I_{Quality}$ and $I_{Quantity}$ are distributed independently of each other by construction.

In column (4), we assess the potential for confounding due to unobserved differences across individuals. In spite of the rich control set $\mathbf{X}_i(a)$, our results may still reflect genetic nurture effects and selection effects due to unobservables. We follow Oster (2019) and calculate bias-adjusted treatment effects to account for this issue. The procedure assumes that changes in the coefficients of interest due to the introduction of $\mathbf{X}_i(a)$ are informative for the extent of confounding due to unobservables. The estimator requires two key inputs. The first input is R^2_{max}—the R^2 from a hypothetical regression of educational attainment on our variables of interest as well as observed and unobserved controls. The second input is δ—a measure for the relative degree of confounding through observed and unobserved controls. We follow the suggestion of Oster (2019) and specify R^2_{max} as 1.3 times the empirical R^2 from column

[14] In comparison to Figure 5.2, there are minor changes in coefficients due to the correlation of PGS^{EA} and $I_{Quality}$. This correlation, however, is small and does not threaten the identification of the gene-environment interaction—see our discussion in section 5.5.1.

(3), and $\delta = 1$. Intuitively, $\delta = 1$ assumes that observed and unobserved confounders are equally related to the treatment.[15] The results remain qualitatively unaltered in comparison to column (3), yet the point estimates of PGS^{EA} and $I_{Quality}$ drop significantly. For example, under the maintained assumptions of R^2_{max} and $\delta = 1$, the point estimate of PGS^{EA} drops by almost half in comparison to our baseline estimate. However, note that it is also $\approx 50\%$ lower than in a sibling fixed effect estimation (Appendix Table A5.2). The latter controls perfectly for genetic nurture effects. Hence, in view of our rich controls for family socio-economic background, the assumption of $\delta = 1$ is likely too conservative.[16]

In Appendix Figure A5.2, we provide further sensitivity analyses with respect to the influence of unobserved confounding variables. Following the procedure of Cinelli and Hazlett (2020), we show that our results for PGS^{EA} and its interaction with $I_{Quality}$ remain statistically significant at the 5%-level even if the set of unobserved confounding variables were more than ten times as strong as paternal education. Similarly, our results for $I_{Quality}$ would remain statistically significant at the 5%-level if the set of unobserved confounding variables were more than five times as strong as paternal education. In view of the strong correlation of parental education with genetic endowments, its decisive role for school choices, and its strong predictive power for educational outcomes of children, these results bestow further confidence into the fact that our results are not just a reflection of genetic nurture effects and selection into schools by family background.

In principle, the negative gene-environment interaction between PGS^{EA} and $I_{Quality}$ could be driven by low PGS^{EA} students gaining from higher-quality teachers, or high PGS^{EA} students losing from higher-quality teachers. In Figure 5.4, we provide evidence for the former, but not the latter. In this figure, we show years of education as predicted from the estimates in column (3) of Table 5.2. Moving horizontally from left to right at a given PGS^{EA} level, we see that predicted education increases strongly in the lower parts of the PGS^{EA} distribution. To the contrary, in the upper parts of the PGS^{EA} distribution, predicted education remains virtually unchanged regardless of the quality of teachers at a given school. This pattern is encouraging as it suggests that investments into teacher quality mitigate inequity in educational outcomes without compromising the attainment of genetically advantaged students.

[15] Cinelli and Hazlett (2020) question this interpretation of δ as it is a function of (i) the association of the *treatment variable* with observed and unobserved confounders and (ii) the association of the *outcome variable* with observed and unobserved confounders. Therefore, Cinelli and Hazlett (2020) propose a bounding procedure based on parameter k_D that varies with (i) but not with (ii). Implementing their alternative procedure, our main conclusions remain unaffected—see our discussion in the following.

[16] Another popular way of reporting the results from sensitivity analyses à la Oster (2019) is to calculate the level of δ required to make coefficients equal to zero. We report these levels and associated bootstrapped standard errors for PGS^{EA}, $I_{Quality}$, and their interaction in the following: PGS^{EA} (1.989 [0.358]), $I_{Quality}$ (1.263 [0.503]), $PGS^{EA} \times I_{Quality}$ (-7.731 [-37.409]). Note that $PGS^{EA} \times I_{Quality}$ is very insensitive to the inclusion of controls. Therefore, standard errors are large and the corresponding point estimate for δ cannot be reliably calculated.

5.5.3 Robustness Analysis

We probe the robustness of our results in two steps. First, we investigate whether $I_{Quality}$ and $I_{Quantity}$ pick up the effect of other school characteristics that may correlate with student outcomes. Second, we test whether our estimates of the complementarity parameter κ are confounded by interactions between genetic endowments and family environments.

Other school characteristics. First, in Figure 5.2 we document positive sorting into schools based on $I_{Quality}$. As a consequence, students in schools with high-quality teachers may additionally be exposed to a more favorable composition of their peer group. A broad literature has documented that skill formation is influenced by school peers (Sacerdote, 2014; Bietenbeck, 2019; Isphording and Zölitz, 2020).[17] Hence, our results for $I_{Quality}$ may reflect both the quality of teachers and peer group composition. To test this hypothesis, we make use of Add Health's in-school questionnaire that elicits background information from a total of 90,000 students in the sampled schools. Based on this questionnaire, we calculate proxy indicators for the quality of the student's peers. In particular, we calculate (i) average years of paternal education, (ii) the share of single parent families, and (iii) student's average self-assessment with respect to the likelihood of attaining a college degree.[18] Then, we include these indicators as well as their interaction with PGS^{EA} into our estimation model.

Table 5.3 displays the results. Column (1) replicates our baseline estimates. In columns (2)–(4), we sequentially introduce the peer quality indicators as well as their interaction with PGS^{EA}. Each proxy for the quality of peers is highly predictive of educational attainment. For example, a 1 SD increase in average paternal education of peers is associated with a 0.26 increase in years of education. Importantly, however, for all considered peer quality indicators our conclusions with respect to $I_{Quality}$, $I_{Quantity}$, and their interaction with genetic endowments remain unaffected.

Second, $I_{Quality}$ and $I_{Quantity}$ may be correlated with school rules and sanction policies. Existing literature suggests that school rules may promote educational attainment (Bacher-Hicks et al., 2019). For example, the success of charter schools has been attributed to strict "no excuses" policies (Angrist et al., 2013). Hence, our results for $I_{Quality}$ may reflect both the quality of teachers and the effect of school rules. To test this hypothesis, we exploit information from headmaster questionnaires and conduct a principal component analysis on various school

[17] Sotoudeh et al. (2019) show genetic endowments of peers are also associated with individual outcomes.

[18] To avoid mechanical relationships between individual characteristics and peer group composition we calculate leave-one-out (jackknife) indicators. A detailed description of these variables is disclosed in Supplementary Material A5.3.

policies.[19] We extract three components that reflect the school's strictness regarding (i) drug use, (ii) social misconduct, and (iii) academic misconduct.

In columns (5)–(7) of Table 5.3, we sequentially introduce the strictness indicators as well as their interaction with PGS^{EA}. Neither of the indicators is predictive of educational attainment, nor is there an interaction with genetic endowments. Our conclusions with respect to $I_{Quality}$, $I_{Quantity}$, and their interaction with genetic endowments remain unaffected.

Third, there may be unobservable school characteristics that drive the relationship between $I_{Quality}$, $I_{Quantity}$, and educational attainment. To address this concern, we use transcript records across grades 9–12 of roughly 12,000 Add Health respondents to calculate cohort-specific measures of school value-added in GPAs for Science, Math, and English. In the extant literature, value-added measures are mostly calculated with respect to test scores that are unaffected by evaluation biases of teachers. To the contrary, GPAs capture student progress in cognitive and behavioral outcomes as well as teacher perceptions (Jackson, 2018). In spite of these intricacies, GPAs are highly predictive of long-term student outcomes (Borghans et al., 2016; Kirkebøen, 2021). Therefore, GPA-based value-added measures provide a good way to capture the quality of schooling environments beyond the measures reported headmaster surveys and administrative data. Specifically, we follow the indirect calculation procedures suggested in Chetty et al. (2014a) and Jackson et al. (2020): we residualize subject-specific GPAs from lagged GPAs in English, Math, and Science, lagged and contemporaneous measures of tracks in these subjects as well as a rich set of individual background characteristics. In turn, we sum residuals to calculate school-times-cohort fixed effects. To avoid mechanical relationships between individual outcomes and cohort-specific school effects, we calculate leave-cohort-out predictions while giving greater weight to neighboring cohorts. We calculate these measures separately for each subject but summarize the school-specific information by extracting the first principal component from the three value-added measures (see Supplementary Material A5.3 for details).

In column (8) of Table 5.3, we introduce school value-added as well as its interaction with PGS^{EA} as additional controls. While school value-added is indeed predictive of educational attainment, there is no effect heterogeneity across the PGS^{EA} distribution. Furthermore, the associations of $I_{Quality}$, $I_{Quantity}$, and PGS^{EA} with educational attainment remain unaffected. Hence, we find no evidence that our relationships of interest are confounded by unobservable school characteristics.

Family environments. In our baseline analysis, we control for a rich set of parental background characteristics to control for genetic nurture effects and selection into schools. How-

[19] In wave 1, headmasters were asked about the school's policy in the following domains of behavior: cheating, fighting with or injuring another student, alcohol or drug possession, drinking alcohol or using illegal drugs, smoking, verbally or physically abusing a teacher, and stealing school property. Possible policies are (i) no policy, (ii) verbal warning, (iii) minor action, (iv) in-school suspension, (v) out-of-school suspension, and (vi) expulsion. A detailed description of these variables is disclosed in Supplementary Material A5.3.

ever, even if we were able to perfectly control for these confounding factors, the complementarity parameter for genetic endowments and school investments may still be confounded by interactions between genetic endowments and family socio-economic status (Keller, 2014; Domingue et al., 2020). To test this hypothesis, we enrich our estimation model by interacting genetic endowments and school environments with the entire control vector $\mathbf{X}_i(a)$. Therefore, we allow for the possibility that family socio-economic status interacts with both genetic endowments and school investments.

Table 5.4 displays the results. Column (1) replicates our baseline estimates. Column (2) displays the enriched estimation model. In spite of a slight decrease in precision, our conclusions with respect to $I_{Quality}$, $I_{Quantity}$, and their interaction with PGS^{EA} remain unaffected.

In addition, we assess whether $I_{Quality}$, $I_{Quantity}$ and their interaction with genetic endowments predict parental investments.[20] A positive association of PGS^{EA} and $I_{Quality}$ with parental investments may suggest that $\mathbf{X}_i(a)$ does not fully account for genetic nurture effects and selection into schools. Therefore, we collect information on a series of activities that the child has conducted with her mother or father in the last four weeks.[21] Following Kling et al. (2007) and Anderson (2008) we standardize each response dimension to $\mu = 0$ and $\sigma = 1$, and sum them linearly by parent to obtain aggregate indexes of time investment. We then use resulting indexes for parental time investment as the outcome of interest, respectively.

The results are shown in columns (3)–(4) of Table 5.4. PGS^{EA} and parental investments are indeed positively associated. However, this association does not necessarily imply the existence of genetic nurture effects. Instead they could also reflect *evocative gene-environment correlations*, i.e. that children select into environments depending on their genetic endowments G_i (Smith-Woolley et al., 2018). If we were to follow this interpretation, the observed association is not indicative for a third factor confounding the relationship of interest, but rather speaks to a particular mechanism of how genetic endowments influence child outcomes. Importantly, there is no differential association between PGS^{EA} and parental time investments depending on school quality.

Lastly, in column (5) of Table 5.4, we again estimate a fully interacted but also incorporate controls for family environments that are potentially endogenous to PGS^{EA} and schooling environments. In particular, we include the indexes for maternal and paternal time investments as well as the log of annual family income. Despite a decrease in sample size and the associated loss in precision, our results remain unaffected.

[20] Note that parental investments may reflect responses to PGS^{EA}, $I_{Quality}$, and $I_{Quantity}$. Therefore, we analyze them as separate outcomes instead of including them in $\mathbf{X}_i(a)$.

[21] These activities include shopping, playing sports, church attendance, talking about dates, going to movies and similar events, talking about personal problems, having an argument, talking about school work, working together on school work, and talking about other things at school. See Supplementary Material A5.3 for details.

Overall, these results bolster confidence that our estimates for the complementarity parameter κ are not confounded by interactions between genetic endowments and family environments.

5.5.4 Mechanisms

In this section, we analyze mechanisms that underpin the substitutability of genetic endowments and teacher quality. We abstract from $I_{Quantity}$ in view of its robust non-association with educational outcomes (see sections 5.5.1–5.5.3).

Educational Degrees. Total years of education summarizes information from various educational stages, where each stage requires a different mix of skills θ_i (Cunha et al., 2006, 2010). Therefore, we repeat our analysis by replacing total years of education with binary variables for whether respondents achieved (i) at least a high school degree or GED, (ii) completed a 2-year college degree, (iii) completed a 4-year college degree, and (iv) completed a post-graduate degree.

In Figure 5.5 we display the resulting point estimates for the complementarity parameter κ and the associated 95% confidence bands.

The series in circles indicates that the compensating effect of teacher quality has a U-shaped pattern throughout the educational life-cycle. There is a small reduction of the probability to drop out of high school, followed by larger substitutability with respect to 2-year and 4-year college degrees. The substitutability of high-quality teachers and genetic endowments levels off at the post-graduate level. This pattern is consistent with the following interpretation. High school graduation is a relatively "inclusive" educational outcome that is accessible for most, including low PGS^{EA} students in low-quality schooling environments. Evidence to this effect is provided by a high school graduation rate of 97% in our sample (Table 5.1). To the contrary, post-graduate education is a relatively "exclusive" educational outcome that is accessible for only few students, and where a certain level of genetic endowments *and* conducive environments is relevant. In both cases, there is limited scope for high-quality teachers to make a difference for low PGS^{EA} students. College education, however, takes a middle ground between these two polar outcomes and therefore offers scope for disadvantageous genetic endowments to be offset by conducive school environments and vice versa.[22]

The series in triangles indicates analogous complementarity parameters for genetic endowments and a summary index of family socio-economic status (SES).[23] Consistent with Pa-

[22] Gene-environment interactions on binary outcomes may be misinterpreted when estimated in a linear probability model. As a remedy Domingue et al. (2020) recommend to present results for the underlying continuous variable, i.e. years of education in the case of this paper. Reassuringly, our results for years of education are in line with our results for educational degrees.

[23] In particular we use the "social origins score" from Belsky et al. (2018) measured at wave 1. Results for alternative measures of family SES such as family income or potential wages of either parent are similar and available upon request.

pageorge and Thom (2020) and Buser et al. (2021a), the complementarity between genetic endowments and family SES increases across the educational life-cycle of individuals. The differential complementarity patterns of school investment and family SES point to the complexity of the skill production function where endowments and different investments interact in distinct and time-variant ways across the life-cycle of individuals.

Skill Formation. In section 5.3, we formulated educational attainment Y_i as a function of child skills θ_i at the end of childhood. Skills that influence educational attainment are multidimensional and comprise a broad set of health indicators and (non-)cognitive skills (Almlund et al., 2011; Heckman and Mosso, 2014). Furthermore, an emerging literature provides evidence for each of these skill dimensions being partially shaped by genetic influence (Demange et al., 2020; Buser et al., 2021a; Demange et al., 2021).

We evaluate these potential channels by analyzing the associations of PGS^{EA} and $I_{Quality}$ with a set of intermediate outcomes. In terms of health outcomes, we focus on subjective health as measured by quality-adjusted life years (QALY) and objective health as measured by an index that comprises information about whether the respondent is obese, afflicted by stage one hypertension, or has high cholesterol. In terms of cognitive skills, we use the Picture Vocabulary Test (PVT) as a measure for verbal intelligence. Lastly, we focus on personality and preferences as two distinct conceptualizations of non-cognitive skills (Becker et al., 2012; Humphries and Kosse, 2017). In particular, we focus on risk aversion and patience, and the Big Five personality traits. All measures are collected in waves 3 and 4 of Add Health, and hence after respondents had left high school but potentially before they had concluded their highest level of education (see Supplementary Material A5.3 for details).

Health, cognitive skills, risk aversion, and patience have been shown to be strong predictors of educational attainment (Jackson, 2009; Burks et al., 2015; Castillo et al., 2018a,b). Furthermore, openness and emotional stability—the inverse of neuroticism—are positively associated with educational attainment (Becker et al., 2012; Buser et al., 2021b). Based on this evidence, we expect positives association of both PGS^{EA} and $I_{Quality}$ with each of these intermediate outcomes. The sign of the gene-environment interaction is a priori unclear. However, in view of the substitutability of PGS^{EA} and $I_{Quality}$ in the production of educational attainment, we expect similar substitutability patterns for a subset of these intermediate outcomes as well.

Table 5.5 summarizes the results. In column (1)–(2) of Panel (a), we focus on health outcomes. In line with expectations, our results show a positive association of PGS^{EA} with both subjective and objective health. A 1 SD increase in PGS^{EA} increases subjective (objective) health by 0.069 SD (0.043 SD). Furthermore, the negative coefficient on the interaction of PGS^{EA} and $I_{Quality}$ suggests that this increase is particularly pronounced for low PGS^{EA} students: a 1 SD increase in teacher quality reduces the positive association of subjective health with the PGS^{EA} by $\approx 41\%$ ($= 0.028/0.069$).

In column (3) of Panel (a), we focus on the PVT as a measure of cognitive skills. In line with expectations, our results show positive associations of both PGS^{EA} and $I_{Quality}$ with the PVT. A 1 SD increase in PGS^{EA} ($I_{Quality}$) is associated with a 0.181 SD (0.102 SD) increase in the PVT. Furthermore, both factors are substitutes for each other. A 1 SD increase in teacher quality reduces the positive association of PVT and PGS^{EA} by $\approx 19\%$ ($= 0.034/0.181$).

In columns (4)–(5) of Panel (a), we focus on economic preferences. In line with expectations, we find strong positive associations of PGS^{EA} with both risk aversion and patience. A 1 SD increase in PGS^{EA} is associated with a 0.038 SD (0.074 SD) increase in risk aversion (patience). Furthermore, PGS^{EA} and $I_{Quality}$ are substitutes for each other. A 1 SD increase in $I_{Quality}$ reduces the positive associations of risk aversion and patience with the PGS^{EA} by $\approx 124\%$ ($= 0.046/0.037$) and $\approx 59\%$ ($= 0.044/0.074$), respectively.

In Panel (b), we focus on personality traits. In line with expectations, we find a positive association of PGS^{EA} with openness and a negative association of PGS^{EA} with neuroticism. However, $I_{Quality}$ is not predictive of any of the Big Five dimensions. Furthermore, we do not find evidence for complementarity of PGS^{EA} and $I_{Quality}$ in the production of personality traits.

To summarize: we find negative gene-environment interactions of genetic endowments and teacher quality in the production of subjective health, cognitive skills, risk aversion and patience. Given their predictive power for educational attainment, these intermediate outcomes are plausible channels to explain the substitutability of genetic endowments and teacher quality in the production of educational attainment.

5.6 Conclusion

The question of how natural endowments and environmental factors determine life outcomes has a long history of inquiry in philosophy and science (Descartes, 1641; Locke, 1690; Lamarck, 1838; Darwin, 1859). The assumption that life outcomes are the result of genetic and environmental factors initially led to the so-called "nature versus nurture" debate. However, current research has moved beyond this simplistic dichotomy and recognizes that individual life outcomes are the result of a complex interplay between nature and nurture. Importantly, this insight illustrates that the importance of genetic endowments for life outcomes is not immutable. Instead, it opens an avenue for policy interventions that shape the relevant environment.

In this paper, we contribute to this research agenda by studying the interplay of genetic endowments and schooling environments in the production of educational outcomes. Making use of recent advances in molecular genetics, we link an individual-level index of genetic predispositions for educational success with measures of our environmental factors of interest, namely teacher "quality" and "quantity" during high school. In turn, we can investigate directly

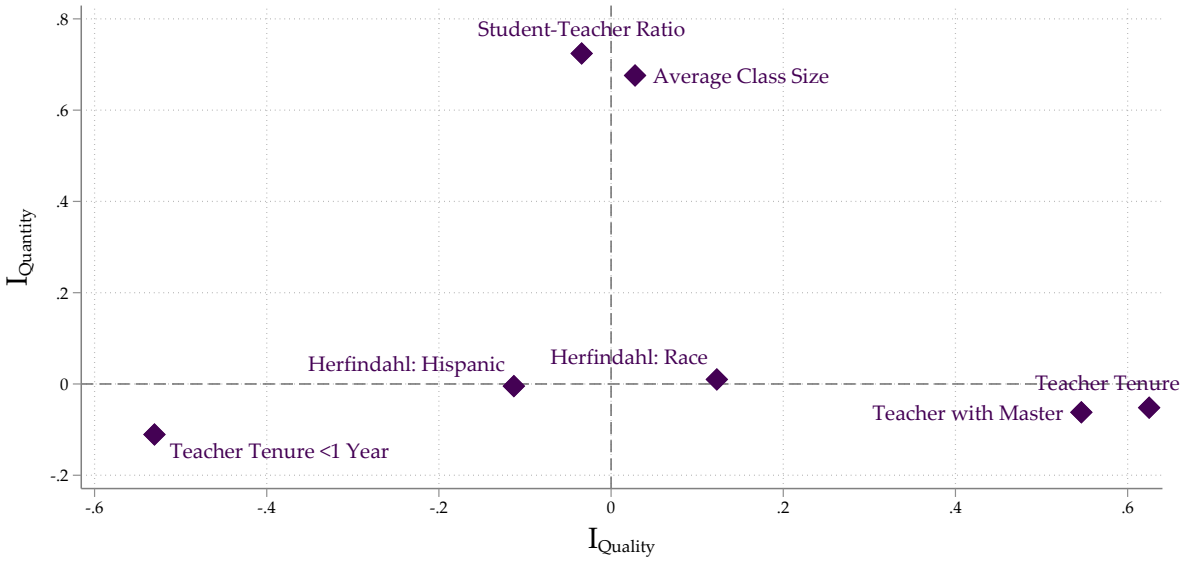
whether the importance of genetic endowments varies by the quality of their high school environments.

Our findings suggest that school investments have the potential to cushion the genetic gradient in educational attainment. However, this conclusion depends on the particular type of investment. On the one hand, increases in "teacher quality" offset genetic disadvantages. On the other hand, we do not find any substitutability with respect to "teacher quantity." Our findings furthermore suggest that increased gains in educational attainment for students of low genetic endowments are mediated by gains in subjective health, cognitive skills and risk aversion, and patience.

Genes are important co-determinants of many life outcomes, including education. However, although they are fixed at conception, their importance can be mediated by suitable policy intervention. In the case of education, increasing the quality of teachers in high schools may provide an important step to leveling the playing field for all students regardless of their draw in the genetic lottery.

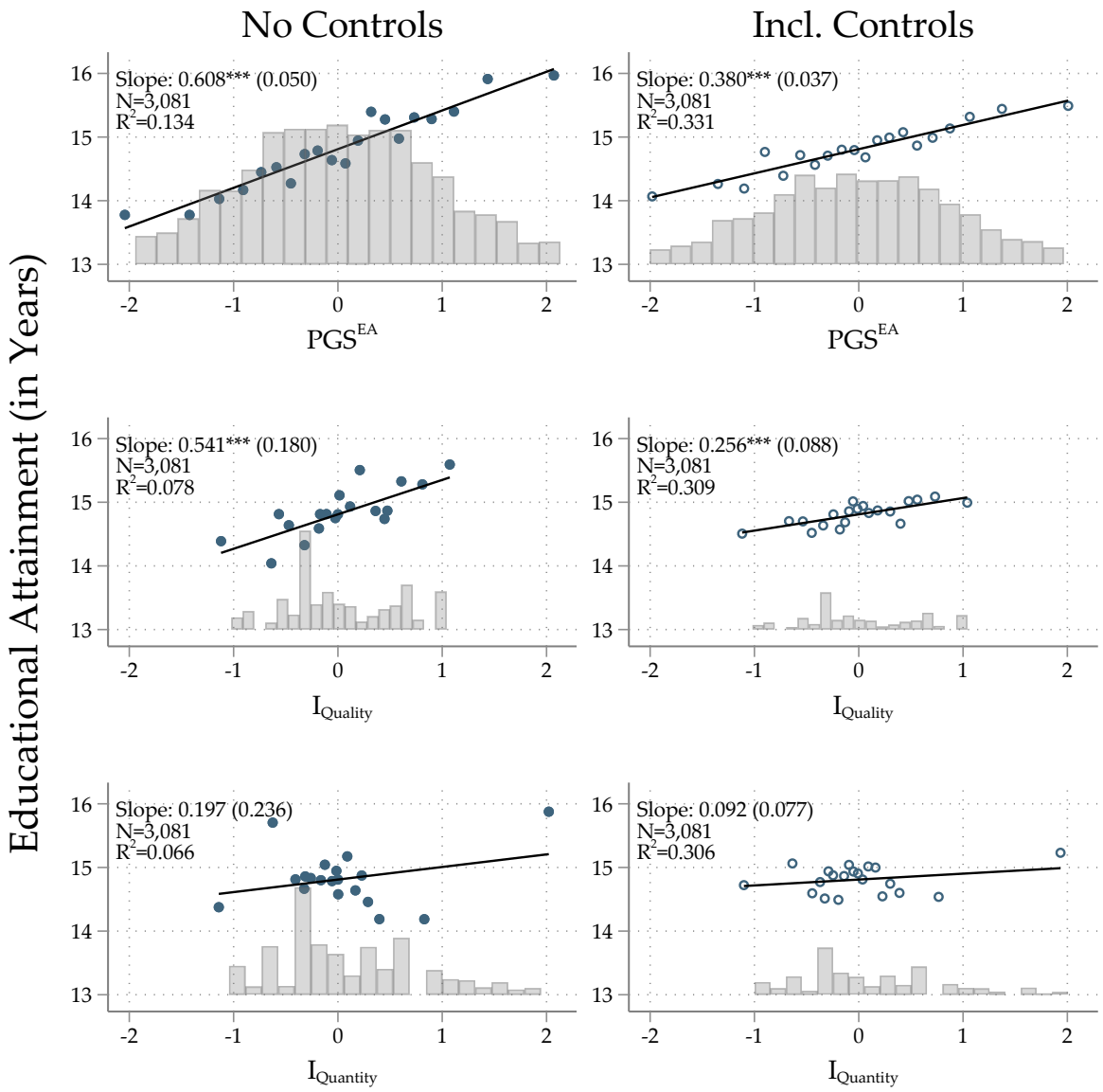
Figures and Tables

Figure 5.1: Rotated loadings on factors for school characteristics



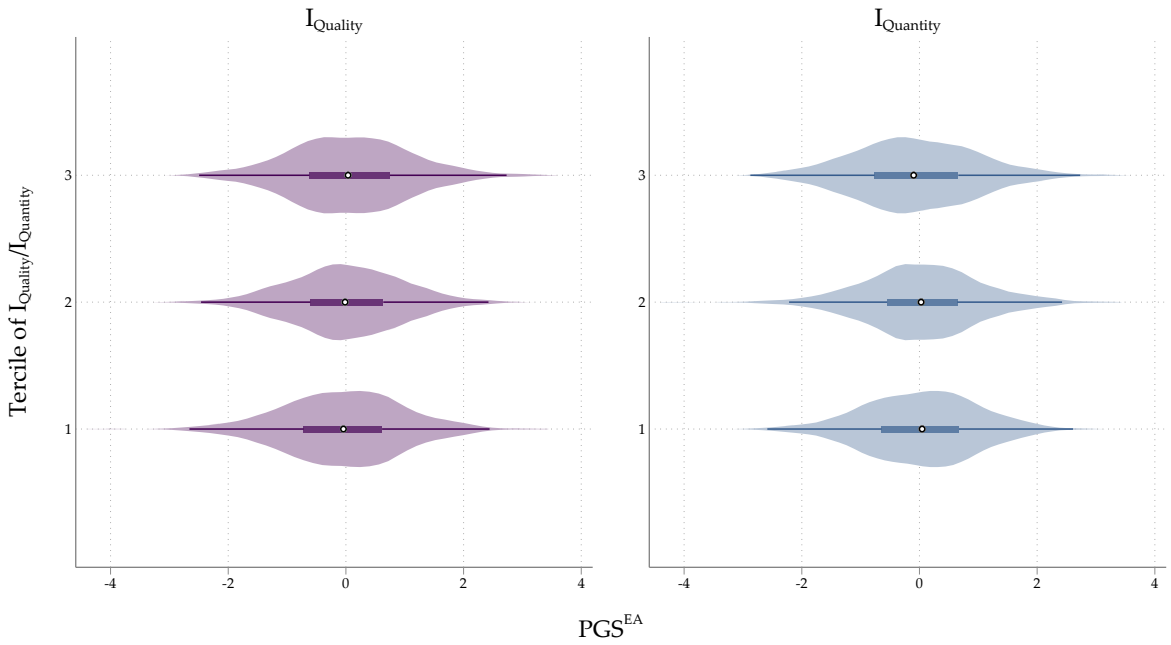
Notes: Own Calculations. This figure shows the rotated factor loadings on $I_{Quality}$ and $I_{Quantity}$. The principal component analysis is conducted using the following school-level information: (i) average class size, (ii) average student-teacher ratio, (iii) share of teachers with a Master degree, (iv) share of new teachers in the current school year, (v) share of teachers with school-specific tenure of more than five years, and Herfindahl indices to measure teacher diversity with respect to (vi) race and (vii) Hispanic background. Data source: National Longitudinal Study of Adolescent to Adult Health.

Figure 5.2: Association of educational attainment with PGS^{EA}, $I_{Quality}$, and $I_{Quantity}$



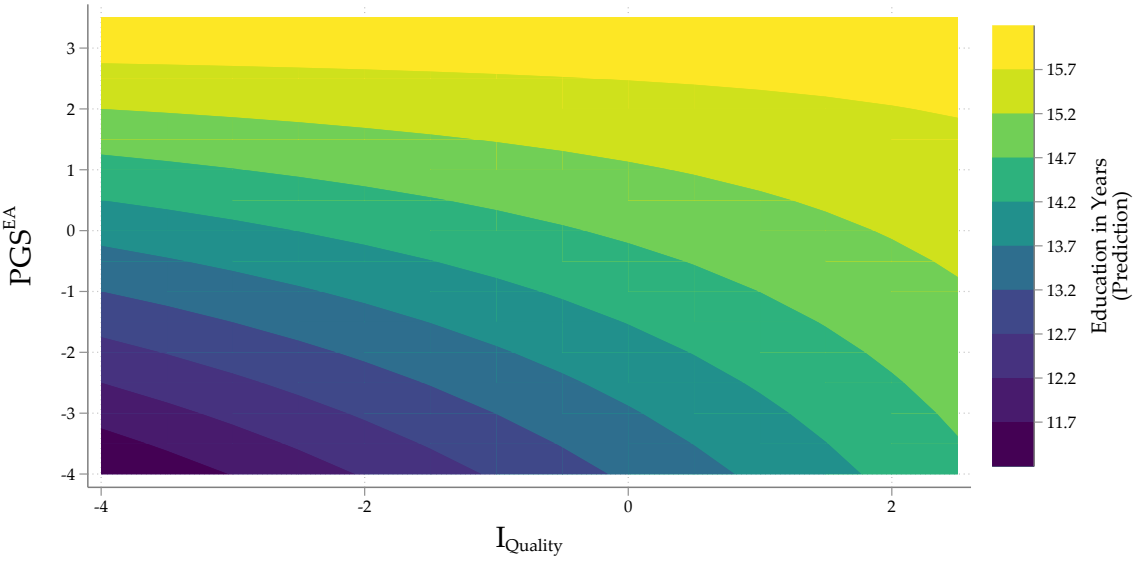
Notes: Own Calculations. This figure visualizes the correlation of completed years of education with PGS^{EA}, $I_{Quality}$, and $I_{Quantity}$, respectively. We bin scatterplots using 20 quantiles of the variable of interest. Gray bars indicate density distributions of the (residualized) variable of interest. Black lines are fitted from linear regressions of educational attainment on the variable of interest. In the left-column we control for state fixed effects, only. In the right column we introduce the full set of control variables. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Figure 5.3: PGS^{EA} distribution by $I_{Quality}$ and $I_{Quantity}$



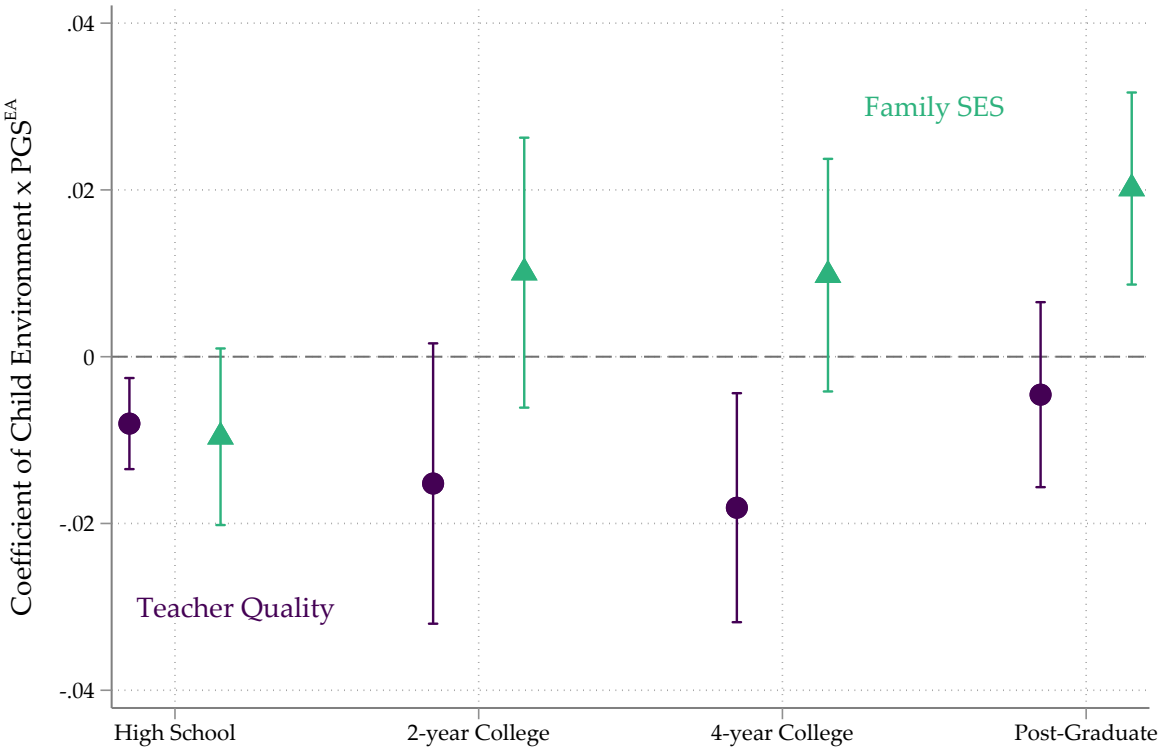
Notes: Own Calculations. This figure shows unconditional PGS^{EA} distribution by terciles of both $I_{Quality}$ and $I_{Quantity}$. The central point indicates the median, the bar indicates the interquartile range. The density is represents the estimated Epanechnikov kernel density. Data: National Longitudinal Study of Adolescent to Adult Health

Figure 5.4: Association of PGS^{EA} with years of education by $I_{Quality}$



Notes: Own Calculations. This figure shows a prediction of completed years of education by PGS^{EA} and $I_{Quality}$ cell. Predictions are calculated from the model estimated in column (3) of Table 5.2. Data source: National Longitudinal Study of Adolescent to Adult Health

Figure 5.5: Association of PGS^{EA} and school/family environments with degree attainment



Notes: Own Calculations. This figure shows point estimates and 95% confidence bands of interaction associations between PGS^{EA} and school/family environments during childhood with completed education degrees. Estimates follow the specification of equation (5.7). *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. Standard errors are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Table 5.1: Summary statistics

	N=3,081; Siblings=525; High Schools=77			
	Mean	SD	Min	Max
Educational Attainment				
Years Education	14.81	2.25	8.00	20.00
High School Degree	0.97	0.18	0.00	1.00
2-year College Degree	0.53	0.50	0.00	1.00
4-year College Degree	0.42	0.49	0.00	1.00
Post-Graduate Degree	0.15	0.36	0.00	1.00
Variables of Interest				
PGS^{EA}	0.00	1.00	-4.18	3.35
$I_{Quality}$	0.00	1.00	-3.41	1.91
$I_{Quantity}$	0.00	1.00	-3.25	3.21
Child Background Characteristics				
Female	0.55	0.50	0.00	1.00
Age in Months (Wave 1)	193.64	19.76	144.00	256.00
Maternal Age at Birth	25.49	4.83	16.00	44.33
Christian	0.82	0.38	0.00	1.00
Education Mother (in Years)	13.63	2.50	8.00	19.00
Education Father (in Years)	13.67	2.68	8.00	19.00
Potential Wage/Hour Mother	12.61	1.38	9.45	14.27
Potential Wage/Hour Father	15.48	1.31	11.14	17.11

Notes: Own Calculations. This table shows summary statistics for the core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion. Data source: National Longitudinal Study of Adolescent to Adult Health

Table 5.2: Association of PGS^{EA} and school environments with years of education

	Baseline			Oster (2019)
Outcome: Years of Education	(1)	(2)	(3)	(4)
PGS^{EA}	0.374*** (0.033)	0.376*** (0.037)	0.371*** (0.033)	0.202*** (0.044)
$I_{Quality}$	0.227*** (0.081)	–	0.222*** (0.083)	0.047 (0.078)
$PGS^{EA} \times I_{Quality}$	-0.073** (0.033)	–	-0.072** (0.033)	-0.082** (0.035)
$I_{Quantity}$	–	0.064 (0.068)	0.062 (0.058)	-0.012 (0.066)
$PGS^{EA} \times I_{Quantity}$	–	0.036 (0.035)	0.026 (0.031)	-0.031 (0.040)
Child Controls	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓
N	3,081	3,081	3,081	3,081
R^2	0.335	0.332	0.335	–
R^2_{max}	–	–	–	0.436
Outcome Mean	14.810	14.810	14.810	14.810
Outcome SD	2.250	2.250	2.250	2.250

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. The first panel establishes our baseline estimates. The second panel displays bias-adjusted treatment effects following the procedure of Oster (2019). We impose R^2_{max} by multiplying R^2 from column (3) with 1.3. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Standard errors of bias adjusted treatment effects are bootstrapped based on 200 draws. Data source: National Longitudinal Study of Adolescent to Adult Health

Table 5.3: Robustness to additional school characteristics

	Baseline	+ School Peer Characteristics			+ School Sanction Policies			+ School VA (GPA)
Outcome: Years of Education	(1)	Educ. Father (2)	Single Parents (3)	College Aspir. (4)	Drugs (5)	Social (6)	Acad. (7)	(8)
PGS^{EA}	0.371***	0.361***	0.370***	0.360***	0.370***	0.368***	0.368***	0.367***
	(0.033)	(0.035)	(0.035)	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)
$I_{Quality}$	0.222***	0.150**	0.221***	0.211***	0.209**	0.195**	0.231**	0.171**
	(0.083)	(0.072)	(0.082)	(0.073)	(0.091)	(0.088)	(0.093)	(0.085)
$PGS^{EA} \times I_{Quality}$	-0.072**	-0.077**	-0.073**	-0.076**	-0.072**	-0.073**	-0.069**	-0.071**
	(0.033)	(0.036)	(0.033)	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)
$I_{Quantity}$	0.062	-0.033	-0.011	-0.017	0.055	0.038	0.030	0.014
	(0.058)	(0.059)	(0.056)	(0.056)	(0.063)	(0.057)	(0.061)	(0.068)
$PGS^{EA} \times I_{Quantity}$	0.026	0.017	0.030	0.024	0.027	0.026	0.022	0.030
	(0.031)	(0.028)	(0.032)	(0.030)	(0.032)	(0.032)	(0.032)	(0.034)
School Characteristic	–	0.261***	-0.201***	0.212***	0.003	-0.108	0.071	0.109*
		(0.055)	(0.049)	(0.045)	(0.052)	(0.076)	(0.072)	(0.062)
$PGS^{EA} \times$ School Characteristic	–	-0.044	0.031	-0.046	0.002	0.024	0.029	-0.016
		(0.042)	(0.039)	(0.035)	(0.041)	(0.030)	(0.039)	(0.031)
Child Controls	✓	✓	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓	✓	✓
N	3,081	2,965	2,965	2,965	2,999	2,999	2,999	2,773
R^2	0.335	0.344	0.343	0.343	0.338	0.339	0.338	0.315

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. We control for additional school characteristics and their interaction with PGS^{EA}. The relevant school characteristics are indicated in the column header. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Table 5.4: Robustness to family environments

	Outcome: Years of Education		Outcome: Parental Investment		Outcome: Years of Education
	Baseline	Full Interaction	Mother	Father	Endogenous Controls
	(1)	(2)	(3)	(4)	(5)
PGS^{EA}	0.371***	0.414***	0.044**	0.058***	0.339***
	(0.033)	(0.086)	(0.017)	(0.017)	(0.099)
$I_{Quality}$	0.222***	0.198*	0.004	0.034	0.219*
	(0.083)	(0.116)	(0.043)	(0.042)	(0.126)
$PGS^{EA} \times I_{Quality}$	-0.072**	-0.092***	-0.013	-0.009	-0.098**
	(0.033)	(0.035)	(0.014)	(0.016)	(0.040)
$I_{Quantity}$	0.062	0.064	0.015	-0.013	0.061
	(0.058)	(0.093)	(0.040)	(0.036)	(0.112)
$PGS^{EA} \times I_{Quantity}$	0.026	0.008	-0.019	-0.031*	0.014
	(0.031)	(0.032)	(0.018)	(0.016)	(0.037)
Child Controls	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓
Full Interaction	×	✓	×	×	✓
Endogenous Controls	×	×	×	×	✓
N	3,081	3,081	3,081	2,541	2,125
R^2	0.335	0.354	0.101	0.078	0.379

Notes: Own Calculations. The first panel of this table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. In column (2) we control for all possible interactions between PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ and the control variables. The second panel of this table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with an index of parental time investments. The third panel of this table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education while accounting for endogenous control variables. Endogenous control variables include the index for maternal time investments, the index for paternal time investments, and log family income. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Table 5.5: Association of PGS^{EA} and school environments with skill measures

Panel (a)	Health		Cognitive	Preferences	
	Subjective (1)	Objective (2)	PVT (3)	Risk (4)	Patience (5)
PGS^{EA}	0.069***	0.043**	0.181***	0.038**	0.074***
	(0.017)	(0.018)	(0.017)	(0.015)	(0.017)
$I_{Quality}$	0.021	0.032	0.102***	0.047	0.045
	(0.042)	(0.036)	(0.039)	(0.030)	(0.038)
$PGS^{EA} \times I_{Quality}$	-0.028**	-0.000	-0.034*	-0.046***	-0.044***
	(0.014)	(0.020)	(0.018)	(0.015)	(0.013)
Child Controls	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓
N	3,081	3,081	3,001	3,077	3,077
R^2	0.078	0.054	0.207	0.112	0.096

Panel (b)	Personality				
	Open-ness (1)	Conscient-iousness (2)	Extra-version (3)	Agree-ableness (4)	Neuro-ticism (5)
PGS^{EA}	0.073***	-0.017	-0.006	0.038*	-0.084***
	(0.017)	(0.017)	(0.019)	(0.020)	(0.019)
$I_{Quality}$	0.038	-0.031	-0.043	0.057	-0.018
	(0.033)	(0.036)	(0.030)	(0.037)	(0.033)
$PGS^{EA} \times I_{Quality}$	0.012	-0.007	-0.001	-0.007	0.023
	(0.013)	(0.015)	(0.023)	(0.019)	(0.019)
Child Controls	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓
N	3,059	3,079	3,075	3,077	3,077
R^2	0.084	0.041	0.031	0.133	0.092

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with health, cognitive skills, preferences, and personality. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Appendix

A5.1 Supplementary Tables

Table A5.1: Sample representativeness

	Population (Cohorts 1974-1983)		Analysis Sample	
	All	Non-Hispanic White	Unweighted	Re-Weighted
Gender				
Male	0.498	0.503	0.453	0.503
Female	0.502	0.497	0.547	0.497
Education Mother				
≤ High School	0.536	0.489	0.494	0.489
> High School; < College Degree	0.281	0.302	0.217	0.301
≥ College Degree	0.183	0.209	0.289	0.210
Education Father				
≤ High School	0.472	0.425	0.491	0.425
> High School; < College Degree	0.255	0.271	0.196	0.271
≥ College Degree	0.273	0.304	0.312	0.303
Age Mother at Birth				
< 25 Years	0.353	0.330	0.485	0.330
≥ 25 Years	0.647	0.670	0.515	0.670
Parental Income				
< $50,000	0.557	0.491	0.531	0.516
≥ $50,000; < $100,000	0.352	0.403	0.390	0.401
≥ $100,000	0.091	0.106	0.079	0.083
Education Respondent				
≤ High School	0.301	0.225	0.181	0.173
> High School; < College Degree	0.327	0.344	0.399	0.402
≥ College Degree	0.372	0.431	0.419	0.425

Notes: Own Calculations. This table shows summary statistics of the core analysis sample in comparison to other population samples. It shows respondents' characteristics for the following samples: (i) the US population from birth cohorts 1974–1983, (ii) the Non-Hispanic White US population from birth cohorts 1974–1983, (iii) the core estimation sample, and (iv) the core estimation sample re-weighted to match (ii) with respect to *Gender*, *Education Mother*, *Education Father*, and *Age Mother at Birth*. Population data on *Gender* and *Education Respondent* from IPUMS ACS 2019 (Ruggles et al., 2020). Population data on *Education Mother*, *Education Father*, *Age Mother at Birth*, and *Parental Income* from IPUMS CPS 1994 (Flood et al., 2020). Data source: National Longitudinal Study of Adolescent to Adult Health, American Community Survey (ACS), Current Population Survey (CPS)

Table A5.2: Association of PGS and years of education: Comparing between-family and within-family models

	Between-Family	Within-Family
Outcome: Years of Education	(1)	(2)
PGS^{EA}	0.374^{***} (0.037)	0.458^{***} (0.160)
$I_{Quality}$	0.226^{***} (0.083)	–
$I_{Quantity}$	0.063 (0.057)	–
Child Controls	✓	✓
Family Controls	✓	✓
Sibling Fixed Effect	×	✓
N	$3,081$	525
R^2	0.334	0.785
Outcome Mean	14.810	14.928
Outcome SD	2.250	2.262

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. The first panel establishes our baseline estimates based on a between-family model. The second panel displays results from a family fixed effect model using within-family variation only. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

Table A5.3: Tests for equality of PGS^{EA} distributions

	Terciles of $I_{Quality}$ / $I_{Quantity}$		
	1	2	3
Panel (a): $I_{Quality}$			
1	–	–	–
2	0.66	–	–
3	0.25	0.77	–
Panel (b): $I_{Quantity}$			
1	–	–	–
2	0.77	–	–
3	1.00	0.70	–

Notes: Own Calculations. This table shows the results of pairwise Kolmogorov-Smirnov tests for the PGS^{EA} distributions within different terciles of $I_{Quality}$ and $I_{Quantity}$, respectively. Results are summarized by the p-value for the null hypothesis that the two PGS^{EA} distributions are equal within the corresponding terciles of $I_{Quality}$ and $I_{Quantity}$. Data source: National Longitudinal Study of Adolescent to Adult Health

Table A5.4: Alternative polygenic scores

	Baseline	+ Controls for Other Polygenic Scores					
Outcome: Years of Education	(1)	Body Mass Index (2)	ADHD (3)	Depressive Symptoms (4)	Intelligence (5)	Ever Smoker (6)	Sleep Duration (7)
PGS^{EA}	0.371***	0.357***	0.346***	0.372***	0.358***	0.347***	0.374***
	(0.033)	(0.035)	(0.032)	(0.034)	(0.039)	(0.038)	(0.033)
$I_{Quality}$	0.222***	0.227***	0.223***	0.226***	0.228***	0.225***	0.227***
	(0.083)	(0.082)	(0.081)	(0.081)	(0.081)	(0.080)	(0.081)
$PGS^{EA} \times I_{Quality}$	-0.072**	-0.084**	-0.074**	-0.073**	-0.079**	-0.080**	-0.073**
	(0.033)	(0.036)	(0.033)	(0.034)	(0.037)	(0.037)	(0.033)
Other PGS	–	-0.070**	-0.130***	-0.024	0.029	-0.128***	-0.004
		(0.032)	(0.034)	(0.032)	(0.040)	(0.041)	(0.032)
Other PGS $\times$ $I_{Quality}$	–	-0.037	0.019	0.007	0.011	-0.019	0.003
		(0.037)	(0.031)	(0.030)	(0.040)	(0.039)	(0.034)
Child Controls	✓	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓	✓
N	3,081	3,081	3,081	3,081	3,081	3,081	3,081
R^2	0.335	0.336	0.338	0.335	0.335	0.338	0.335

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. We control for other PGS and their interaction with $I_{Quality}$ and $I_{Quantity}$. The relevant PGS are indicated in the column header. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

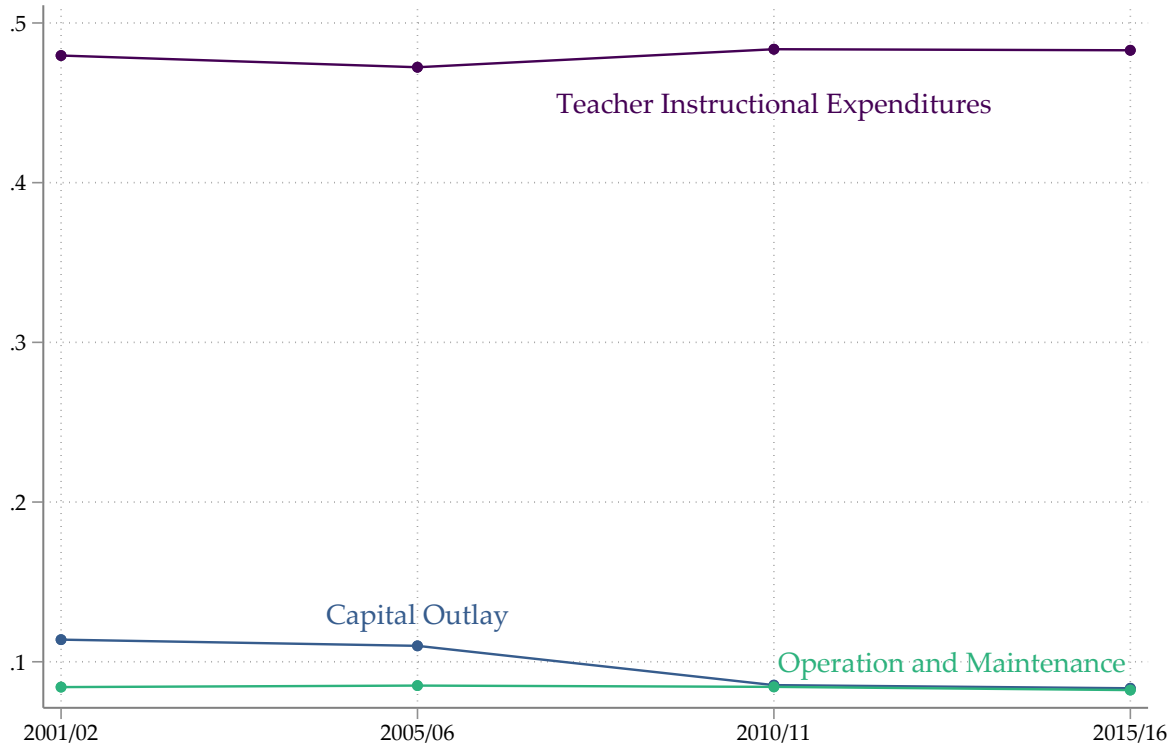
Table A5.5: Robustness to sample composition

	Baseline	Alternative Sample Composition		
Outcome: Years of Education	(1)	Re-Weighted (2)	Excl. all (Potential) Movers (3)	Inc. all (Potential) Movers (4)
PGS^{EA}	0.371***	0.359***	0.360***	0.381***
	(0.033)	(0.035)	(0.040)	(0.028)
$I_{Quality}$	0.222***	0.198**	0.154	0.217***
	(0.083)	(0.084)	(0.099)	(0.078)
$PGS^{EA} \times I_{Quality}$	-0.072**	-0.073**	-0.068*	-0.050*
	(0.033)	(0.034)	(0.039)	(0.029)
$I_{Quantity}$	0.062	0.049	0.067	0.050
	(0.058)	(0.062)	(0.069)	(0.075)
$PGS^{EA} \times I_{Quantity}$	0.026	0.037	0.013	0.037
	(0.031)	(0.034)	(0.035)	(0.023)
Child Controls	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓
N	3,081	3,027	2,526	4,185
R^2	0.335	0.315	0.328	0.319

Notes: Own Calculations. This table shows the joint association of PGS^{EA}, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. In column (2) we re-weight our analysis sample to match ACS and CPS with respect to gender composition, educational attainment of parents, and the age of mothers at birth—see also Appendix Table A5.1. In column (3) we exclude respondents that visit feeder schools in wave 1 and for whom we do not have information on subsequent high schools. In column (4) we include respondents that are in Add Health high schools in wave 1 and for whom we do not have information on subsequent high schools. *Child Controls*: Gender times birth cohort dummies, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, a dummy for Christian religion, state fixed effects. All non-binary variables are standardized on the estimation sample to have $\mu = 0, \sigma = 1$. Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

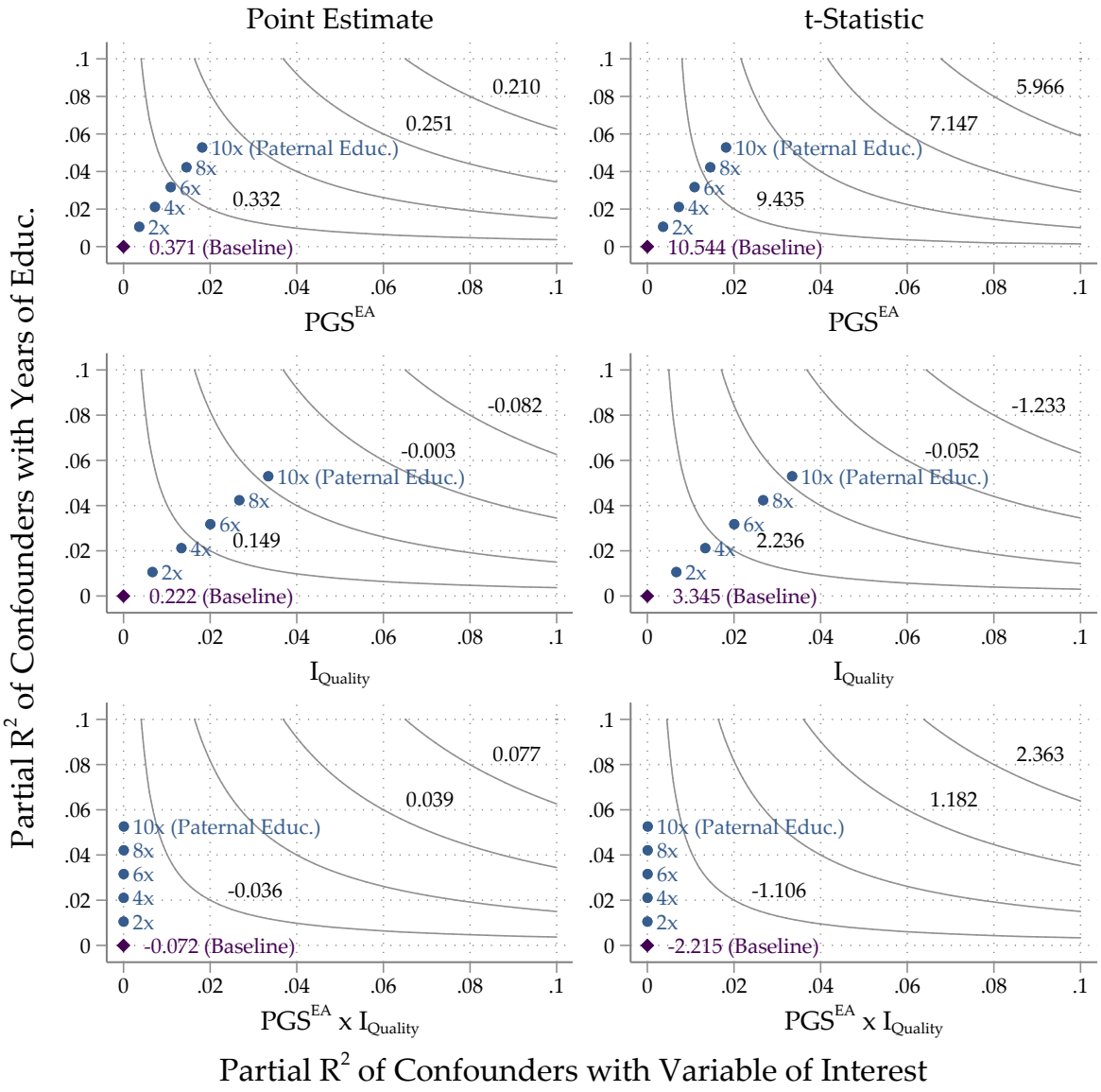
A5.2 Supplementary Figures

Figure A5.1: Top 3 school expenditure categories (in % of total)



Notes: Own Calculations. This figure shows per-pupil expenditures shares in public elementary and secondary schools in the US. Teacher Instructional Expenditures includes teachers' salaries and employee benefits. Capital Outlay includes expenditures for property and for buildings and alterations completed by school district staff or contractors. Operation and Maintenance includes expenditures for the supervision of operations and maintenance, the operation of buildings, the care and upkeep of grounds and equipment, vehicle operations (other than student transportation) and maintenance, and security. Data source: Common Core of Data (CCD), National Public Education Financial Survey

Figure A5.2: Sensitivity to unobserved confounders



Notes: Own Calculations. This figure shows the sensitivity of the point estimates for PGS^{EA}, $I_{Quality}$, and their interaction to unobserved confounding variables. Following the procedure of Cinelli and Hazlett (2020), we calculate the bias-adjusted treatment effect of PGS^{EA} and $I_{Quality}$, and their interaction under different assumptions about the partial R^2 of confounding variables with the variables of interest and the partial R^2 of confounding variables with years of education. Each contour line shows point estimates (left-hand panel) and t-statistics (right-hand panel) for different combinations of the two partial R^2. Each circle shows resulting values for different multiples of confounders as strong as parental education. Diamonds show baseline estimates from Table 5.2. Standard errors are clustered at the school level. Data source: National Longitudinal Study of Adolescent to Adult Health

A5.3 Data Appendix

Outcome Variables

Educational Attainment. We measure educational attainment by total *years of education*. In each wave, respondents were asked about their highest level of education at the time of the interview. For each respondent, we use the most recent information and transform education levels into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: eighth grade or less (8), some high school (10), high school graduate (12), GED (12), some vocational/technical training (13), some community college (14), some college (14), completed vocational/technical training (14), associate or junior college degree (14), completed college (16), some graduate school (17), completed a master's degree (18), some postbaccalaureate professional education (18), some graduate training beyond a master's degree (19), completed post-baccalaureate professional education (19), completed a doctoral degree (20).

We use the most recent available information to construct the following measures for educational degrees: *High School* (including GED), *2-year College*, *4-year College*, and *Post-Graduate*. Two-year college degrees include associate and junior college degrees as well as vocational and technical training after high school. Four-year college degrees include bachelor's degrees. Post-graduate degrees include master's degrees, doctoral degrees, and post-baccalaureate professional degrees. If available, information is taken from wave 5; otherwise we take it from waves 4 or 3, respectively. We only include respondents for which we observe educational degrees when they are at least 27 years old at the time of observation. We assume an ordinal ranking of degrees (high school < 2-year college < 4-year college < post-graduate) and assign the possession of a lower-ranked degree if a respondent obtained a higher-ranked degree. For example, we assume that a respondent has finished high school if he or she has obtained a college degree, even if we don't have explicit information about high school graduation status.

Health. We proxy *subjective health* by quality-adjusted life years (QALY) that we derive from self-assessed health (SAH) measures. We use information from waves 3 and 4, where participants were asked "in general, how is your health?" We convert their (categorical) responses into a continuous measure using a mapping proposed by Van Doorslaer and Jones (2003). Using information about objective health—the Health Utility Index Mark III—Van Doorslaer and Jones (2003) scale the intervals of the SAH categories. This approach yields "quality weights" for health between 0 and 1. The values for each health status category are as follows (quality weights in parentheses): "excellent" (0.9833), "very good" (0.9311), "good" (0.841), "fair" (0.707), and "poor" (0.401).[1] We average resulting QALY measures across waves 3 and 4.

We construct an index of *objective health* based on information from wave 4. Specifically, we sum the standardized values about whether a respondent (i) is obese, (ii) has stage one

[1] See Table 4 in Van Doorslaer and Jones (2003).

hypertension, and (iii) has high cholesterol (as indicated by the respondent). Each item was answered with either "yes" (= 1) or "no" (= 0). We reverse-code our measure of objective health such that higher values indicate better health.

Cognitive Skills. The *Picture Vocabulary Test* (PVT) is a test for receptive hearing vocabulary and is a widely-used proxy for verbal ability and scholastic aptitude. To administer the PVT, an examiner presents a series of pictures to the respondent. There are four pictures per page, and the examiner speaks a word describing one of the pictures. The respondent then has to indicate the picture that the word describes. In our analysis we use age-adjusted PVT percentile ranks from wave 3 (Harris, 2020).

Preferences. We construct two measures of preferences: *risk aversion* and *patience*. In waves 3 and 4, participants were asked (i) whether they like to take risks, and (ii) whether they live their life without much thought for the future. Questions were answered on a five-point Likert scale ranging from "strongly agree" to "strongly disagree." We reverse-code both measures and use averages from waves 3 and 4 in our analysis.

Personality. The Big Five personality traits are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Almlund et al., 2011). We use information from wave 4 to construct personality measures. Participants were asked a set of questions that each relate to one of the five personality traits. Questions were answered on a five-point Likert scale ranging from "strongly agree" to "strongly disagree." We use averages of the following questions in our analysis. *Openness*: (i) "I have a vivid imagination," (ii) "I have difficulty understanding abstract ideas" (reverse-coded), (iii) "I am not interested in abstract ideas" (reverse-coded), (iv) "I do not have a good imagination" (reverse-coded). *Conscientiousness*: (i) "I get chores done right away," (ii) "I like order," (iii) "I often forget to put things back in their proper place" (reverse-coded), (iv) "I make a mess of things" (reverse-coded). *Extraversion*: (i) "I am the life of the party," (ii) "I talk to a lot of different people at parties," (iii) "I don't talk a lot" (reverse-coded), (iv) "I keep in the background" (reverse-coded). Agreeableness: (i) "I sympathize with others' feelings," (ii) "I feel others' emotions," (iii) "I am not interested in other people's problems" (reverse-coded), (iv) "I am not really interested in others" (reverse-coded). *Neuroticism*: (i) "I have frequent mood swings," (ii) "I get upset easily," (iii) "I am relaxed most of the time" (reverse-coded), (iv) "I seldom feel blue" (reverse-coded).

Parental Investment. To measure *parental time investments*, we use information on a series of activities that children have done with their mother or father in the last four weeks. Specifically, the child is asked whether he or she has (i) gone shopping, (ii) played a sport, (iii) gone to a religious service or church-related event, (iv) talked about someone he or she is dating, or a party he or she went to, (v) gone to a movie, play, museum, concert, or sports event, (vi) had a talk about a personal problem he or she was having, (vii) had a serious argument about him or her behavior, (viii) talked about his or her school work or grades, (ix) worked on a project for school, (x) talked about other things he or she is doing in school. Questions were answered

with "yes" (= 1) or "no" (= 0). We standardize each response to have mean zero and standard deviation one and then sum by parent (Kling et al., 2007; Anderson, 2008).

Variables of Interest

Polygenic Scores. Add Health obtained saliva samples from consenting participants in wave 4. After quality control procedures, genotyped data is available for 9,974 individuals and 609,130 SNPs. Add Health uses this data and calculates a set of different PGS using summary statistics from existing GWAS. Our baseline measure PGS^{EA} is based on statistics from Lee et al. (2018). In our analysis, we also use the PGS for body mass index (*BMI*) (Yengo et al., 2018), attention deficit hyperactivity disorder (*ADHD*) (Demontis et al., 2019), *depressive symptoms* (Howard et al., 2019), *intelligence* (Savage et al., 2018), *smoking* (Liu et al., 2019), and *sleep duration* (Jansen et al., 2019). All polygenic scores are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of genotyped Add Health respondents.

School Characteristics. In wave 1 and 2, Add Health administered questionnaires to headmasters of Add Health schools. We use this information to construct indicators for high school investments using a principal components analysis that includes the following school-level information: (i) average class size, (ii) share of teachers with a master degree, (iii) share of new teachers in the current school year, (iv) share of teachers with school-specific tenure of more than five years, and Herfindahl indices to measure teacher diversity with respect to (v) race and (vi) Hispanic background.[2] We also include school-level information about the average student-teacher ratio (number of full-time students per full-time equivalent teachers) in 1995/96 taken from the Common Core of Data (CCD) and the Private School Survey (PSS). We apply a factor rotation for interpretability reasons (oblique oblimin rotation of the Kaiser normalized matrix with $\gamma = 0$; see Gorsuch, 1983). The first component loads almost exclusively on average class size and average student-teacher ratio. Hence, we interpret this component, $I_{Quantity}$, as an indicator for the "quantity" of teachers. The second component primarily loads positively on the percentage of teachers with a master degree and the share of teachers with a tenure of more than five years; it loads negatively on the share of new teachers in the current school year. We interpret this component, $I_{Quality}$, as an indicator for the "quality" of teachers. Both factors are coded such that higher values indicate higher school investments, i.e. higher teacher "quantity" investments (smaller classes) and higher teacher "quality" investments (better teachers), respectively. The calculated factors are orthogonal to each other by construction. They are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health high schools.[3]

[2] Herfindahl indices are calculated by first squaring the share of each component and then summing up resulting values (i.e. $H = \sum_{i=1}^{N} a_i^2$, where a_i is the share of component i, and N is the total number of components). For the Herfindahl index for race, we include the schools' share of full-time classroom teachers that are (i) White, (ii) Black or African American, (iii) American Indian or Native American, (iv) and Asian or Pacific Islander. For the Herfindahl index for Hispanic background, we include the schools' share of full-time classroom teachers that are (i) Hispanic or of Spanish origin, and (ii) neither Hispanic nor of Spanish origin.

[3] Note that in an oblique rotation, factors may be slightly correlated.

Table A5.6: Summary statistics (outcomes)

	Obs.	Mean	SD	Min	Max
Educational Attainment					
Years Education	3,081	14.81	2.25	8.00	20.00
High School Degree	3,081	0.97	0.18	0.00	1.00
2-year College Degree	3,081	0.53	0.50	0.00	1.00
4-year College Degree	3,081	0.42	0.49	0.00	1.00
Post-Graduate Degree	3,081	0.15	0.36	0.00	1.00
Health					
Subjective	3,081	0.91	0.07	0.40	0.98
Objective	3,081	0.03	1.94	-6.46	1.62
Cognitive Skills					
Picture Vocabulary Test	3,001	59.94	25.94	0.00	100.00
Preferences					
Risk Aversion	3,077	2.83	0.86	1.00	5.00
Patience	3,077	3.93	0.72	1.00	5.00
Personality					
Openness	3,059	3.63	0.63	1.00	5.00
Conscientousness	3,079	3.65	0.70	1.25	5.00
Extraversion	3,075	3.33	0.77	1.00	5.00
Agreeableness	3,077	3.87	0.58	1.00	5.00
Neuroticism	3,077	2.56	0.70	1.00	5.00
Parental Time Investments					
Mother	3,081	0.53	4.34	-8.51	14.89
Father	2,541	0.32	4.28	-6.47	16.74

Notes: Own Calculations. This table shows summary statistics for outcome variables in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion. Data source: National Longitudinal Study of Adolescent to Adult Health

Family Socio-Economic Status. We use the *social origins factor score* constructed by Belsky et al. (2018). Their measure uses information about parental education, parental occupation, household income, and household receipt of public assistance in wave 1. The score is standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health respondents in wave 1.

Table A5.7: Summary statistics (variables of interest)

	Obs.	Mean	SD	Min	Max
Polygenic Scores					
PGS^{EA}	3,081	0.05	1.00	-4.13	3.39
BMI	3,081	-0.02	1.01	-3.42	3.56
ADHD	3,081	-0.05	1.00	-3.82	3.48
Depressive Symptoms	3,081	-0.02	1.01	-3.79	3.55
Intelligence	3,081	0.02	0.99	-3.30	4.06
Ever Smoker	3,081	-0.04	1.00	-4.25	4.25
Sleep Duration	3,081	0.02	0.99	-3.74	2.99
School Characteristics					
$I_{Quality}$	3,081	0.07	1.17	-3.90	2.30
$I_{Quantity}$	3,081	-0.03	1.02	-3.34	3.25
Family SES					
Social Origins Factor Score	3,024	0.37	1.12	-4.40	3.51

Notes: Own Calculations. This table shows summary statistics for variables of interest in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion. Data source: National Longitudinal Study of Adolescent to Adult Health

Control Variables

Child Characteristics. The child's *gender* (female or male, as indicated by the interviewer) is taken from the in-home questionnaire in wave 1.

We calculate the child's *age* (in months) at each wave by subtracting the child's birth date from the date of interview. Because birth dates have minor inconsistencies across waves, we take averages across waves 1 to 4.

We use the first 20 *principal components* of full matrix of the genetic data. The components are obtained from a principal components analysis on the matrix of SNPs in Add Health (see

Braudt and Harris, 2020, for a discussion). The principal components are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of genotyped Add Health respondents.

Family Socio-Economic Status. We use information from wave 1 to construct measures of *parents' education*. We transform parents' highest degree into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: never went to school (0), eighth grade or less (8), some high school (10), completed vocational/technical training instead of high school (10), went to school but level unknown (12), respondent doesn't know (12), high school graduate (12), GED (12), completed vocational/technical training after high school (14), some college (14), completed college (16), professional training beyond a master's degree (19). Where available, mothers' and fathers' education refers to the resident parent. If this information is not available, we use the biological parents' education instead.

Information about *mother's age at birth* (in years) is obtained from wave 1 if available, and wave 2 otherwise. To calculate age at birth, we take information about mother's age (as indicated by the child) and subtract the age of the child at the respective wave.

Information about religion (*Christian* or not) is obtained from wave 1 (as indicated by the child).

We calculate *potential wages* for population group g in time period t according to the following formula (Shenhav, 2021):

$$\hat{w}_{gt} = \sum_{j} \frac{E_{jg,1970}}{E_{g,1970}} \times \sum_{o} \frac{E_{ojg,1970}}{E_{jg,1970}} \left(\pi_{ojt,-r}\right) \times w_{ojt,-r},$$

where $\frac{E_{jg,1970}}{E_{g,1970}}$ describes the group-specific employment share of industry j in 1970, $\frac{E_{ojg,1970}}{E_{jg,1970}}$ describes the group- and industry-specific employment share of occupation o in 1970, $\pi_{ojt,-r}$ describes the leave-region-out industry-specific employment growth in occupation o for year t relative to 1970 (scaled by the overall employment growth in occupation o for year t relative to 1970), and $w_{ojt,-s}$ describes the leave-region-out average hourly wage paid in year t for each occupation/industry/region cell. We define groups g by individuals that are homogeneous in gender (male, female), educational attainment (< High School, High School, > High School), and ethnicity (Non-Hispanic White, Hispanic, Non-Hispanic Black). We define regions r by census regions (North-East, Midwest, South, West). Employment shares in 1970 are taken from the 1970 decennial census. Employment shares and wages in periods t are taken from the March Supplements of the Current Population Survey (CPS) over the time period 1975-2000. We match time series of $\hat{w}_{gt}$ to the parents of respondents in Add Health based on information about g. Then we calculate (i) mean potential wages across respondent ages 0–14, and (ii) the standard deviation in potential wages across respondent ages 0–14.

School Characteristics. We use information about school *peer characteristics* from the in-school questionnaire in wave 1. Specifically, for each school we calculate average years of

education of students' fathers, the share of single parents, and the average subjective likelihood of students to attend college. We transform the father's highest degree into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: never went to school (0), eighth grade or less (8), some high school (10), went to school but level unknown (12), respondent doesn't know (12), high school graduate (12), GED (12), completed vocational/technical training after high school (14), some college (14), completed college (16), professional training beyond a four-year college (19). For college aspiration, students indicate how likely it is that they will graduate from college. Responses range from "no chance" (= 0) to "it will happen" (= 8). We define a student to have college aspiration if his or her response is above "about 50-50" (= 4), and to have no college aspiration otherwise. To prevent mechanical correlation between school peer characteristics and respondent characteristics, we calculate averages and shares while excluding individual respondents (leave-one-out).

We use information from the school administrator questionnaire in wave 1 to construct measures of *sanction policies* by means of a principal components analysis. School administrators were asked what happens to a student who is caught in their school (i) cheating, (ii) fighting with another student, (iii) injuring another student, (iv) possessing alcohol, (v) possessing an illegal drug, (vi) possessing a weapon, (vii) drinking alcohol at school, (viii) using an illegal drug at school, (ix) smoking at school, (x) verbally abusing a teacher, (xi) physically injuring a teacher, and (xii) stealing school property. Responses are "minor action", "in-school suspension", "out-of-school suspension", and "expulsion." Administrators were asked about sanctions in response to both first and second occurrences. We apply a factor rotation for interpretability reasons (oblique oblimin rotation of the Kaiser normalized matrix with $\gamma = 0$; see Gorsuch, 1983). The first three components load on variables reflecting the school's strictness regarding (i) drug use, (ii) social misconduct, and (iii) academic misconduct. The calculated factors are orthogonal to each other by construction. They are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health high schools.[4]

We calculate *value-added measures* with respect to GPAs in subject s for cohort c visiting high school j following a two-step procedure (Chetty et al., 2014a):

$$\text{GPA}^s_{igjc} = \beta^s Z_{igjc} + \text{VA}^s_{jc} + \epsilon^s_{igjc},$$
$$\widehat{\text{VA}}^s_{jc} = \frac{1}{N}\sum_{i \in jc}^{N} (\text{VA}^s_{jc} + \widehat{\epsilon}^s_{igjc}).$$

Z_{igjc} contains grade fixed effects δ_g, lagged GPAs from grade levels $g-1$ for English, Math and Science as well as current and lagged grade- and subject-specific indicators for academic tracks in English, Math and Science (3 levels per grade times subject cell). To avoid mechanical relationships, we predict $\widetilde{\text{VA}}^s_{jc}$ excluding data from cohort c and choosing a weighting vector $\phi^s = [\phi^s_{c-5}, 05_G enes., \phi^s_{c+5}]$ that minimizes the out-of-sample mean-squared error. Hence,

[4] Note that in an oblique rotation, factors may be slightly correlated.

$\widetilde{\text{VA}}^{s}_{jc}$ is our best prediction based on other cohorts of how much school j will increase GPAs in subject s in one year of high school relative to the improvements of similar students at other schools. We calculate $\widetilde{\text{VA}}^{s}_{jc}$ for English, Math and Science. In turn, we run a principal component analysis and use the first principal component as the aggregate measure of school value-added. The principal component is standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of high schools with available transcript data on Add Health respondents.

Table A5.8: Summary statistics (controls)

	Obs.	Mean	SD	Min	Max
Child Characteristics					
Female	3,081	0.55	0.50	0.00	1.00
Age in Months (Wave 1)	3,081	193.64	19.76	144.00	256.00
Principal Component 1	3,081	0.00	0.01	-0.14	0.10
Principal Component 2	3,081	-0.00	0.01	-0.37	0.07
Principal Component 3	3,081	0.00	0.01	-0.10	0.02
Principal Component 4	3,081	0.00	0.01	-0.09	0.65
Principal Component 5	3,081	-0.00	0.01	-0.07	0.18
Principal Component 6	3,081	-0.00	0.01	-0.14	0.19
Principal Component 7	3,081	-0.00	0.01	-0.13	0.33
Principal Component 8	3,081	-0.00	0.01	-0.37	0.08
Principal Component 9	3,081	0.00	0.01	-0.06	0.07
Principal Component 10	3,081	-0.00	0.01	-0.58	0.26
Principal Component 11	3,081	0.00	0.01	-0.25	0.37
Principal Component 12	3,081	0.00	0.01	-0.39	0.18
Principal Component 13	3,081	-0.00	0.01	-0.35	0.18
Principal Component 14	3,081	-0.00	0.01	-0.12	0.23
Principal Component 15	3,081	0.00	0.01	-0.28	0.23
Principal Component 16	3,081	0.00	0.02	-0.15	0.66
Principal Component 17	3,081	-0.00	0.01	-0.50	0.24
Principal Component 18	3,081	-0.00	0.01	-0.29	0.20
Principal Component 19	3,081	0.00	0.01	-0.26	0.46
Principal Component 20	3,081	-0.00	0.01	-0.18	0.27
Family SES					
Education Mother (in Years)	3,081	13.63	2.50	8.00	19.00
Education Father (in Years)	3,081	13.67	2.68	8.00	19.00
Maternal Age at Birth	3,081	25.49	4.83	16.00	44.33
Christian	3,081	0.82	0.38	0.00	1.00
Potential Wage/Hour Mother (Mean)	3,081	12.61	1.38	9.45	14.27
Potential Wage/Hour Father (Mean)	3,081	15.48	1.31	11.14	17.11
Potential Wage/Hour Mother (SD)	3,081	0.36	0.11	0.12	0.51
Potential Wage/Hour Father (SD)	3,081	0.40	0.08	0.20	0.65
School Characteristics					
Peer Characteristics (Educ. Father)	2,965	13.57	1.05	10.90	17.84
Peer Characteristics (Single Parents)	2,965	0.24	0.08	0.00	0.60
Peer Characteristics (College Aspir.)	2,965	0.76	0.08	0.44	1.00
Sanction Policies (Drugs)	2,999	0.15	1.87	-5.71	9.06
Sanction Policies (Social)	2,999	0.25	1.61	-6.30	5.00
Sanction Policies (Acad.)	2,999	0.04	1.22	-3.41	2.38
Value-Added (GPA)	2,773	0.21	1.55	-4.18	4.41

Notes: Own Calculations. This table shows summary statistics for control variables in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion. Data source: National Longitudinal Study of Adolescent to Adult Health

Bibliography

Aaronson, Daniel, Lisa Barrow, and William Sander (2007). "Teachers and Student Achievement in the Chicago Public High Schools". *Journal of Labor Economics* 25(1), 95–135.

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge (2017). "When should you adjust standard errors for clustering?" *National Bureau of Economic Research Working Paper Series*.

Abdellaoui, Abdel, David Hugh-Jones, Loic Yengo, Kathryn E Kemper, Michel G Nivard, Laura Veul, Yan Holtz, Brendan P Zietsch, Timothy M Frayling, Naomi R Wray, et al. (2019). "Genetic Correlates of Social Stratification in Great Britain". *Nature Human Behaviour* 3(12), 1332–1342.

Acemoglu, Daron and David Autor (2011). "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings". Ed. by David Card and Orley Ashenfelter. Vol. 4. Handbook of Labor Economics. Elsevier, 1043–1171.

Achieve Inc. (2013). "Closing the Expectations Gap: 2013 Annual Report on the Alignment of State K-12 Policies and Practice with the Demands of College and Careers". *ERIC Clearinghouse*.

Aghion, Philippe and Peter Howitt (1998). *Endogenous Growth Theory*. Cambridge: The MIT Press.

Akter, Sonia, Jeff Bennett, and Michael B. Ward (2012). "Climate Change Scepticism and Public Support for Mitigation: Evidence from an Australian Choice Experiment". *Global Environmental Change* 22(3), 736–745.

Alesina, Alberto and Nicola Fuchs-Schündeln (2007). "Good-bye Lenin (or Not?): The Effect of Communism on People's Preferences". *American Economic Review* 97(4), 1507–1528.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2020). "The Polarization of Reality". *AEA Papers and Proceedings* 110, 324–28.

Algan, Yann, Daniel Cohen, Eva Davoine, Martial Foucault, and Stefanie Stantcheva (2021). "Trust in scientists in times of pandemic: Panel evidence from 12 countries". *Proceedings of the National Academy of Sciences* 118(40).

Allen, Joseph P., Robert C. Pianta, Anne Gregory, Amori Yee Mikami, and Janetta Lun (2011). "An Interaction-Based Approach to Enhancing Secondary School Instruction and Student Achievement". *Science* 333(6045), 1034–1037.

Allensworth, Elaine, Sarah Cashdollar, and Julia Gwynne (2021). "Improvements in Math Instruction and Student Achievement Through Professional Learning Around the Common Core State Standards in Chicago". *AERA Open* 7, 1–19.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 4. Elsevier, 1–181.

Almond, Douglas and Bhashkar Mazumder (2013). "Fetal Origins and Parental Responses". *Annual Review of Economics* 5 (1), 37–56.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel (2016). "Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects". Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 5. Handbook of the Economics of Education. Elsevier, 305–396.

Altonji, Joseph G., Erica Blom, and Costas Meghir (2012). "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers". *Annual Review of Economics* 4(1), 185–223.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools". *Journal of Political Economy* 113(1), 151–184.

Altonji, Joseph G. and Richard K. Mansfield (2018). "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects". *American Economic Review* 108(10), 2902–46.

Altunas, Öykü (2017). "Turkey's new school year: Jihad in, evolution out". *BBC*. URL: https://www.bbc.com/news/world-europe-41296714 (visited on 07/09/2021).

American Association for the Advancement of Science (2021). "Dialogue Science, Ethics, and Religion". *Thematic Areas*. URL: https://www.aaas.org/programs/dialogue-science-ethics-and-religion/thematic-areas (visited on 05/08/2021).

Anderson, Michael 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, Vincenzo and Xuejuan Su (2019). "The Impact of Schooling Intensity on Student Learning: Evidence from a Quasi-Experiment". *Education Finance and Policy* 14(4), 679–701.

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters (2016). "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice". *Journal of Labor Economics* 34(2), 275–318.

Angrist, Joshua D. and Victor Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement". *Quarterly Journal of Economics* 114(2), 533–575.

Angrist, Joshua D., Victor Lavy, Jetson Leder-Luis, and Adi Shany (2019). "Maimonides' Rule Redux". *American Economic Review: Insights* 1(3), 309–324.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters (2013). "Explaining Charter School Effectiveness". *American Economic Journal: Applied Economics* 5(4), 1–27.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom (2016a). "College Attrition and the Dynamics of Information Revelation". *National Bureau of Economic Research Working Paper Series, No. 22325*.

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz (2016b). "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California". *American Economic Review* 106(3), 525–62.

Arcidiacono, Peter, V. Joseph Hotz, Arnaud Maurel, and Teresa Romano (2020). "Ex Ante Returns and Occupational Choice". *Journal of Political Economy* 128(12), 4475–4522.

Arold, Benjamin W. (2022). "Evolution vs. Creationism in the Classroom: The Lasting Effects of Science Education". *mimeo*.

Arold, Benjamin W. and M. Danish Shakeel (2021). "The Unintended Effects of Common Core State Standards on Non-Targeted Subjects". *PEPG 21-03. Program on Education Policy and Governance Working Papers Series, Harvard University.*

Arold, Benjamin W., Ludger Woessmann, and Larissa Zierow (2022). "Can Schools Change Religious Attitudes? Evidence from German State Reforms of Compulsory Religious Education". *CESifo Working Paper No. 9504*.

Athey, Susan and Guido W. Imbens (2021). "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption". *Journal of Econometrics, forthcoming*.

Bacher-Hicks, Andrew, Stephen B Billings, and David J Deming (2019). "The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime". *National Bureau of Economic Research Working Paper Series* 26257.

Bailey, Michael, Drew M. Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2020). *Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing During the COVID-19 Pandemic*. Tech. rep.

Baker, Andrew, David F. Larcker, and Charles C.Y. Wang (2021). "How Much Should We Trust Staggered Difference-In-Differences Estimates?" *Rock Center for Corporate Governance at Stanford University Working Paper, No. 246*.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo (2019). "Nation-building Through Compulsory Schooling during the Age of Mass Migration". *Economic Journal* 129(617), 62–109.

Barcellos, Silvia H, Leandro S Carvalho, and Patrick Turley (2021). "The Effect of Education on the Relationship between Genetics, Early-life Disadvantages,and Later-life SES". *National Bureau of Economic Research Working Paper Series* 28750.

Barro, Robert J. (1991). "Economic Growth in a Cross Section of Countries". *Quarterly Journal of Economics* 106(2), 407–443.

– (2001). "Human Capital and Growth". *American Economic Review* 91(2), 12–17.

Barro, Robert J. and Jong Wha Lee (2013). "A New Data Set of Educational Attainment in the World, 1950–2010". *Journal of Development Economics* 104, 184–198.

Barro, Robert J. and Rachel M. McCleary (2003). “Religion and Economic Growth across Countries”. *American Sociological Review* 68(5), 760–781.

– (2005). “Which Countries Have State Religions?” *Quarterly Journal of Economics* 120(4), 1331–1370.

Barth, Daniel, Nicholas W Papageorge, and Kevin Thom (2020). “Genetic Endowments and Wealth Inequality”. *Journal of Political Economy* 128(4), 1474–1522.

Bay-Williams, Jennifer (2016). “Common Core Math in the K-8 Classroom: Results from a National Teacher Survey”. *Thomas B. Fordham Institute*.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx (2020). “Islam and the State: Religious Education in the Age of Mass Schooling”. *National Bureau of Economic Research Working Paper Series No. 27073*.

Beale, H. K. (1941). *A History of the Freedom of Teaching in American Schools in Report of the Commission on the Social Studies, Part 16*. Chicago, IL: Charles Scribner & Sons.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse (2012). “The Relationship Between Economic Preferences and Psychological Personality Measures”. *Annual Review of Economics* 4(1), 453–478.

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

Becker, Sascha O., Markus Nagler, and Ludger Woessmann (2017). “Education and Religious Participation: City-Level Evidence from Germany’s Secularization Period 1890–1930”. *Journal of Economic Growth* 22(3), 273–311.

Becker, Sascha O., Jared Rubin, and Ludger Woessmann (2021). “Religion in economic history: A survey”. *The Handbook of Historical Economics*. Ed. by Alberto Bisin and Giovanni Federico. London: Academic Press, 585–639.

Becker, Sascha O. and Ludger Woessmann (2009). “Was Weber Wrong? A Human Capital Theory of Protestant Economic History”. *Quarterly Journal of Economics* 124(2), 531–596.

– (2013). “Not the Opium of the People: Income and Secularization in a Panel of Prussian Counties”. *American Economic Review, Papers and Proceedings* 103(3), 539–544.

– (2018). “Social Cohesion, Religious Beliefs, and the Effect of Protestantism on Suicide”. *Review of Economics and Statistics* 100(3), 377–391.

Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Loiuse Arseneault, Avshalom Boardman Jason D. amd Caspi, Daton Conley, Jason M. Fletcher, Jeremy Freese, Pamela Herd, Terrie E. Moffitt, Richie Poulton, Kamil Sicinski, Jasmin Wertz, and Kathleen Mullan Harris (2018). “Genetic Analysis of Social-class Mobility in Five Longitudinal Studies”. *Proceedings of the National Academy of Sciences* 115 (31), E7275–E7284.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin W. Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S. Williams, Richie Poulton, and Avshalom Caspi (2016). “The Genetics of Success: How Single-nucleotide

Polymorphisms Associated with Educational Attainment Relate to Life-course Development". *Psychological Science* 27(7), 957–972.

Bentzen, Jeanet Sinding (2019). "Acts of God? Religiosity and Natural Disasters Across Subnational World Districts". *Economic Journal* 129(622), 2295–2321.

Bentzen, Jeanet Sinding and Lena Lindbjerg Sperling (2020). "God politics". CEPR Discussion Paper 14380.

Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann (2021). "Testing". *Journal of Human Resources*.

Berkman, Michael and Eric Plutzer (2011). "Defeating Creationism in the Courtroom, But Not in the Classroom". *Science* 331(6016), 404–405.

Berkman, Michael B., Julianna Sandell Pacheco, and Eric Plutzer (2008). "Evolution and Creationism in America's Classrooms: A National Portrait". *PLoS Biology* 6(5), 920–924.

Betts, Julian R. (1998). "The Impact of Educational Standards on the Level and Distribution of Earnings". *American Economic Review* 88(1), 266–275.

Beuermann, Diether, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo (2018). "What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output". *National Bureau of Economic Research Working Paper Series* 25342.

Biasi, Barbara (forthcoming). "School Finance Equalization Increases Intergenerational Mobility". *Journal of Labor Economics*.

Bietenbeck, Jan (2019). "The Long-term Impacts of Low-achieving Childhood Peers: Evidence from Project STAR". *Journal of the European Economic Association* 18(1), 392–426.

Bishop, John H. (1989). "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review* 79(1), 178–197.

Bishop, John Hillman (1997). "The Effect of National Standards and Curriculum-Based Exams on Achievement". *American Economic Review* 87(2), 260–264.

Bisin, Alberto and Thierry Verdier (2001). "The Economics of Cultural Transmission and the Dynamics of Preferences". *Journal of Economic Theory* 97(2), 298–319.

Blazar, David, Blake Heller, Thomas J Kane, Morgan S. Polikoff, Douglas Staiger, Scott Carrell, Dan Goldhaber, Douglas Harris, Rachel Hitch, Kristian L Holden, et al. (2019). "Learning by the book: Comparing math achievement growth by textbook in six Common Core states". *Center for Education Policy Research, Harvard University*.

Bleiberg, Joshua (2021). "Does the Common Core Have a Common Effect? An Exploration of Effects on Academically Vulnerable Students". *AERA Open* 7, 1–18.

Blossfeld, Hans-Peter and Jutta von Maurice (2011). "Education as a lifelong process: The German National Educational Panel Study (NEPS)". *Zeitschrift für Erziehwissenschaften* 14 [Spezial Issue], 19–34.

Bordon, Paola and Chao Fu (2015). "College-Major Choice to College-Then-Major Choice". *Review of Economic Studies* 82(4), 1247–1288.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries (2016). "What Grades and Achievement Tests Measure". *Proceedings of the National Academy of Sciences* 113(47), 13354–13359.

Borusyak, Kirill, Xavier Jaravel, and Jan Spiess (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". *Working Paper*.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese (2013). "Variation in the Heritability of Educational Attainment: An International Meta-analysis". *Social Forces* 92(1), 109–140.

Braudt, David and Kathleen Mullan Harris (2020). "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health)—Release 2".

Brinch, Christian N. and Taryn Ann Galloway (2012). "Schooling in Adolescence Raises IQ Scores". *Proceedings of the National Academy of Sciences* 109(2), 425–430.

Brown, Sarah and Karl Taylor (2007). "Religion and education: Evidence from the National Child Development Study". *Journal of Economic Behavior and Organization* 63(3), 439–460.

Bryan, Gharad, James J Choi, and Dean Karlan (2021). "Randomizing Religion: the Impact of Protestant Evangelism on Economic Outcomes". *Quarterly Journal of Economics* 136(1), 293–380.

Brzezinski, Adam, Valentin Kecht, David Van Dijcke, and Austin L Wright (2021). "Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States". *Nature Human Behaviour*, 1–9.

Burks, Stephen V., Connor Lewis, Paul A. Kivi, Amanda Wiener, Jon E. Anderson, Lorenz Götte, Colin G. DeYoung, and Aldo Rustichini (2015). "Cognitive skills, personality, and economic preferences in collegiate success". *Journal of Economic Behavior & Organization* 115, 30–44.

Bursztyn, Leonardo, Aakaash Rao, Christopher P. Roth, and David H. Yanagizawa-Drott (2020). "Misinformation During a Pandemic". *National Bureau of Economic Research Working Paper Series, No. 27417*.

Buser, Thomas, Rafael Ahlskog, Magnus Johannesson, Philipp Koellinger, and Sven Oskarsson (2021a). "Using genes to explore the effects of cognitive andnon-cognitive skills on education and labor market outcomes". *Tinbergen Institute Discussion Paper* 2021-088.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2021b). "Can Competitiveness predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures". *National Bureau of Economic Research Working Paper Series* 28916.

Butcher, Kristin F., Patrick J. McEwan, and Akila Weerapana (2014). "The Effects of an Anti-Grade-Inflation Policy at Wellesley College". *Journal of Economic Perspectives* 28(3), 189–204.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021). "Difference-in-Differences with multiple time periods". *Journal of Econometrics* 225(2), 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors". *Review of Economics and Statistics* 90(3), 414–427.

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang (2017). "Curriculum and Ideology". *Journal of Political Economy* 125(2), 338–392.

Cantoni, Davide and Noam Yuchtman (2013). "The Political Economy of Educational Content and Development: Lessons from History". *Journal of Development Economics* 104, 233–244.

Card, David (1999). "Chapter 30 - The Causal Effect of Education on Earnings". Ed. by Orley C. Ashenfelter and David Card. Vol. 3. Handbook of Labor Economics. Elsevier, 1801–1863.

Carmichael, Sheila Byrd, Gabrielle Martino, Kathleen Porter-Magee, and W Stephen Wilson (2010). "The State of State Standards-and the Common Core-in 2010". *Thomas B. Fordham Institute*.

Castillo, Marco, Jeffrey L Jordan, and Ragan Petrie (2018a). "Discount Rates of Children and High School Graduation". *Economic Journal* 129(619), 1153–1181.

– (2018b). "Children's Rationality, Risk Attitudes and Field Behavior". *European Economic Review* 102, 62–81.

CCSSI (2022). *Common Core State Standards Initiative: Standards in Your State*. Accessed on 2022-02-24 at: http://www.corestandards.org/standards-in-your-state.

Cesarini, David and Peter M. Visscher (2017). "Genetics and Educational Attainment". *npj Science of Learning* 2 (4).

Chaisemartin, Clément de and Xavier D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". *American Economic Review* 110(9), 2964–96.

Chanock, Stephen J, Teri Manolio, Michael Boehnke, Eric Boerwinkle, David J Hunter, Gilles Thomas, Joel N Hirschhorn, Gonçalo R Abecasis, David Altshuler, Joan E Bailey-Wilson, et al. (2007). "Replicating Genotype-phenotype Associations". *Nature* (447), 655–660.

Chaudhary, Latika and Jared Rubin (2011). "Reading, writing, and religion: Institutions and human capital formation". *Journal of Comparative Economics* 39(1), 17–33.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star". *Quarterly Journal of Economics* 126(4), 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014a). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates". *American Economic Review* 104 (9), 2593–2632.

– (2014b). "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood". *American Economic Review* 104 (9), 2633–79.

Cinelli, Carlos and Chad Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82(1), 39–67.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2010). "Teacher Credentials and Student Achievement in High School: A Cross-subject Analysis with Student Fixed Effects". *Journal of Human Resources* 45(3), 655–681.

Clots-Figueras, Irma and Paolo Masella (2013). "Education, Language and Identity". *Economic Journal* 123(570), 332–357.

Cohen, Eliot A (2020). "History, Critical and Patriotic: Americans Need a History That Educates but Also Inspires". *Education Next* 20(2), 8–18.

Conger, Dylan, Alec I. Kennedy, Mark C. Long, and Raymond McGhee (2021). "The Effect of Advanced Placement Science on Students' Skills, Confidence, and Stress". *Journal of Human Resources* 56(1), 93–124.

Cortes, Kalena and Joshua Goodman (2014). "Ability-Tracking, Instructional Time, and Better Pedagogy: The Effect of Double-Dose Algebra on Student Achievement". *American Economic Review* 104(5), 400–405.

Costrell, Robert M. (1994). "A Simple Model of Educational Standards". *American Economic Review* 84(4), 956–971.

– (1997). "Can Centralized Educational Standards Raise Welfare?" *Journal of Public Economics* 65(3), 271–293.

Cullen, Julie Berry, Brian A Jacob, and Steven Levitt (2006). "The Effect of School Choice on Participants: Evidence from Randomized Lotteries". *Econometrica* 74(5), 1191–1230.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006). "Interpreting the Evidence on Life Cycle Skill Formation". Ed. by E. Hanushek and F. Welch. Vol. 1. Handbook of the Economics of Education. Amsterdam: Elsevier. Chap. 12, 697–812.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". *Econometrica* 78(3), 883–931.

Cygan-Rehm, Kamila (2021). "Are there no wage returns to compulsory schooling in Germany? A reassessment". *Journal of Applied Econometrics, forthcoming*.

Daro, P., G. B. Hughes, and F. Stancavage (2015). "Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics". *Washington, D.C.: American Institutes for Research*.

Darwin, Charles R. (1859). *On the Origin of Species*. London: John Murray.

Dee, Thomas S. and Brian Jacob (2011). "The impact of no Child Left Behind on student achievement". *Journal of Policy Analysis and Management* 30(3), 418–446.

DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence". *Annual Review of Economics* 2(1), 643–669.

DellaVigna, Stefano and Ethan Kaplan (2007). "The Fox News Effect: Media Bias and Voting". *Quarterly Journal of Economics* 122(3), 1187–1234.

Demange, Perline A., Jouke Jan Hottenga, Abdel Abdellaoui, Margherita Malanchini, Benjamin W. Domingue, Eveline L. de Zeeuw, Kaili Rimfeld, Thalia Eley, Dorret I. Boomsma, Elsje van Bergen, Gerome Breen, Michel G. Nivard, and Rosa Cheesman (2020). "Parental

influences on offspring education: indirect genetic effects of non-cognitive skills". *bioRxiv* 2020.09.15.296236.

Demange, Perline A., Margherita Malanchini, Travis T. Mallard, Pietro Biroli, Simon R. Cox, Andrew D. Grotzinger, Elliot M. Tucker-Drob, Abdel Abdellaoui, Louise Arseneault, Avshalom Caspi, David Corcoran, Benjamin Domingue, Colter Mitchell, Elsje van Bergen, Dorret I. Boomsma, Kathleen M. Harris, Hill F. Ip, Terrie E. Moffitt, Richie Poulton, Joseph Prinz, Karen Sugden, Jasmin Wertz, Benjamin Williams, Eveline L. de Zeeuw, Daniel W. Belsky, K. Paige Harden, and Michel G. Nivard (2021). "Investigating the Genetic Architecture of Non-Cognitive Skills Using GWAS-by-Subtraction". *Nature Genetics* 53, 35–44.

Deming, David J and Kadeem Noray (2020). "Earnings Dynamics, Changing Job Skills, and STEM Careers". *Quarterly Journal of Economics* 135(4), 1965–2005.

Deming, David J. (2017). "The Growing Importance of Social Skills in the Labor Market". *Quarterly Journal of Economics* 132(4), 1593–1640.

Demontis, Ditte, Raymond K Walters, Joanna Martin, Manuel Mattheisen, Thomas D Als, Esben Agerbo, Gısli Baldursson, Rich Belliveau, Jonas Bybjerg-Grauholm, Marie Bækvad-Hansen, et al. (2019). "Discovery of the First Genome-wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder". *Nature Genetics* 51(1), 63–75.

Descartes, René (1641). "Meditations on First Philosophy". *Descartes: Selected Philosophical Writings*. Ed. by John Cottingham. Ed. by Dugald Murdoch. Ed. by Robert Stoothoff. Ed. by Anthony Kenny. Cambridge: Cambridge University Press, 73–122.

Dobzhansky, Theodosius (2013). "Nothing in Biology Makes Sense Except in the Light of Evolution". *American Biology Teacher* 75(2), 87–91.

Domingue, Benjamin W., Daniel Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason Boardman (2015). "Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health". *AERA Open* 1(3), 1–13.

Domingue, Benjamin W., Sam Trejo, Emma Armstrong-Carter, and Elliot M. Tucker-Drob (2020). "Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges". *Sociological Science* 7(19), 465–486.

Dube, Arindrajit, T. William Lester, and Michael Reich (2010). "Minimum wage effects across state borders: estimates using contiguous counties". *Review of Economics and Statistics* 92(4), 945–964.

Dudbridge, Frank (2013). "Power and Predictive Accuracy of Polygenic Risk Scores". *PLOS Genetics* 9 (3), e1003348.

Engelberg, Joseph, Raymond Fisman, Jay C. Hartzell, and Christopher A. Parsons (2016). "Human Capital and the Supply of Religion". *Review of Economics and Statistics* 98(3), 415–427.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya (2011). "Media and Political Persuasion: Evidence from Russia". *American Economic Review* 101(7), 3253–85.

European Commission (2020). “European skills agenda for sustainable competitiveness, social fairness and resilience”. *Publications Office of the European Union Luxembourg*.

Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, David B. Huffman, and Uwe Sunde (2018). “Global Evidence on Economic Preferences”. *Quarterly Journal of Economics* 133(4), 1645–1692.

Feng, Li, David Figlio, and Tim Sass (2018). “School accountability and teacher mobility”. *Journal of Urban Economics* 103, 1–17.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, and J. Robert Warren (2020). “Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]”. *Minneapolis, MN: IPUMS*.

Franck, Raphael and Laurence Iannaccone (2014). “Religious decline in the 20th century West: testing alternative explanations”. *Public Choice* 159(3), 385–414.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek (2013). “Long-term Effects of Class Size”. *Quarterly Journal of Economics* 128(1), 249–285.

Fryer, Roland G. (2011). “Racial inequality in the 21st century: the declining significance of discrimination”. *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4. Elsevier, 855–971.

Fuchs-Schündeln, Nicola and Paulo Masella (2016). “Long-Lasting Effects of Socialist Education”. *Review of Economics and Statistics* 98(3), 428–441.

Gao, Niu and Julien Lafortune (2019). “Common Core State Standards in California: Evaluating Local Implementation and Student Outcomes.” *Public Policy Institute of California*.

GESIS (2019). “German General Social Survey (ALLBUS) – Cumulation 1980-2016.” *GESIS Data Archive ZA4588 Data file version 1.0.0.*

Giuliano, Paola and Antonio Spilimbergo (2014). “Growing up in a Recession”. *Review of Economic Studies* 81(2), 787–817.

Glaeser, Edward L. and Bruce I. Sacerdote (2008). “Education and Religion”. *Journal of Human Capital* 2(2), 188–215.

Goebel, Jan, Markus M Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp (2019). “The German Socio-Economic Panel (SOEP)”. *Journal of Economics and Statistics* 239(2), 345–360.

Goodman, Joshua (2019). “The Labor of Division: Returns to Compulsory High School Math Coursework”. *Journal of Labor Economics* 37(4), 1141–1182.

Goodman-Bacon, Andrew (2021). “Difference-in-Differences with Variation in Treatment Timing”. *Journal of Econometrics, forthcoming*.

Gorsuch, Richard (1983). *Factor Analysis*. 2nd ed. Hillsdale, NJ: Psychology Press.

Gradstein, Mark and Moshe Justman (2002). “Education, Social Cohesion, and Economic Growth”. *American Economic Review* 92(4), 1192–1204.

Graham, Loren (2016). *Lysenko’s Ghost: Epigenetics and Russia*. Harvard University Press.

Grewenig, Elisabeth (2021). “School track decisions and teacher recommendations: Evidence from German state reforms”. *ifo Working Paper No. 353*.

Griliches, Zvi (1992). “The Search for R&D Spillovers”. *Scandinavian Journal of Economics* 94, 29–47.

Gross, Paul R (2005). “The State of State Science Standards”. *Thomas B. Fordham Foundation & Institute*.

Gruber, Jonathan and Daniel M. Hungerman (2008). “The Church Versus the Mall: What Happens When Religion Faces Increased Secular Competition?” *Quarterly Journal of Economics* 123(2), 831–862.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2008). “Social Capital as Good Culture”. *Journal of the European Economic Association* 6, 295–320.

Haberman, Clyde (2017). “Questioning Evolution: The Push to Change Science Class”. *New York Times*. URL: https://www.nytimes.com/2017/11/19/us/retro-report-evolution-science.html (visited on 07/09/2021).

Hahn, Youjin, Liang Choon Wang, and Hee-Seung Yang (2018). “Does greater school autonomy make a difference? Evidence from a randomized natural experiment in South Korea”. *Journal of Public Economics* 161, 15–30.

Hamlin, Daniel and Paul E Peterson (2018). “Have states maintained high expectations for student performance? An analysis of 2017 state proficiency standards”. *Education Next* 18(4), 42–49.

Hanushek, Eric A, Paul E Peterson, and Ludger Woessmann (2012). “Achievement Growth: International and US State Trends in Student Performance”. *Program on Education Policy and Government Report No.: 12-03, Taubman Center for State and Local Government, Harvard Kennedy School*.

Hanushek, Eric A and Steven G Rivkin (2010). “Generalizations about Using Value-added Measures of Teacher Quality”. *American Economic Review* 100(2), 267–271.

Hanushek, Eric A. (1986). “The Economics of Schooling: Production and Efficiency in Public Schools”. *Journal of Economic Literature* 24(3), 1141–1177.

Hanushek, Eric A., Susanne Link, and Ludger Woessmann (2013). “Does school autonomy make sense everywhere? Panel estimates from PISA”. *Journal of Development Economics* 104, 212–232.

Hanushek, Eric A., Steven G. Rivkin, and Jeffrey C. Schiman (2016). “Dynamic Effects of Teacher Turnover on the Quality of Instruction”. *Economics of Education Review* 55, 132–148.

Hanushek, Eric A. and Ludger Woessmann (2006). “Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries”. *Economic Journal* 116(510), C63–C76.

– (2008). “The Role of Cognitive Skills in Economic Development”. *Journal of Economic Literature* 46(3), 607–668.

Hanushek, Eric A. and Ludger Woessmann (2012). “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation”. *Journal of Economic Growth* 17(4), 267–321.

Harris, Kathleen Mullan (2020). “National Longitudinal Study of Adolescent to Adult Health, Public Use Peabody Picture Vocabulary Test (PVT) Score Data, Wave III [dataset]”. *National Longitudinal Study of Adolescent to Adult Health*.

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman (2013). “Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile”. *National Bureau of Economic Research Working Paper Series, No. 19241*.

Havers, Norbert (1972). *Der Religionsunterricht - Analyse eines unbeliebten Fachs*. München: Kösel-Verlag.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz (2010). “Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program”. *Quantitative Economics* 1(1), 1–46.

Heckman, James and Stefano Mosso (2014). “The Economics of Human Development and Social Mobility”. *Annual Review of Economics* 6 (1), 689–733.

Heckman, James J. and Tim Kautz (2012). “Hard evidence on soft skills”. *Labour Economics* 19(4), 451–464.

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2006). “Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond”. Ed. by E. Hanushek and F. Welch. Vol. 1. Handbook of the Economics of Education. Elsevier, 307–458.

Helbig, Marcel and Rita Nikolai (2015). *Die Unvergleichbaren - Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Julius Klinkhardt.

Holden, Kristian L. (2016). “Buy the Book? Evidence on the Effect of Textbook Funding on School-Level Achievement”. *American Economic Journal: Applied Economics* 8(4), 100–127.

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm (2020). “The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills”. *IZA Discussion Paper Series* 13780.

Howard, David M, Mark J Adams, Toni-Kim Clarke, Jonathan D Hafferty, Jude Gibson, Masoud Shirali, Jonathan RI Coleman, Saskia P Hagenaars, Joey Ward, Eleanor M Wigmore, et al. (2019). “Genome-wide Meta-analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions”. *Nature Neuroscience* 22(3), 343–352.

Humphries, John Eric and Fabian Kosse (2017). “On the Interpretation of Non-cognitive Skills—What is Being Measured and Why it Matters”. *Journal of Economic Behavior & Organization* 136, 174–185.

Hungerman, Daniel (2014). “The effect of education on religion: Evidence from compulsory schooling laws”. *Journal of Economic Behavior and Organization* 104(C), 52–63.

Huxley, T. H (1880). “The Coming of Age of the Origin of Species”. *Nature* 22(549), 1–4.

Iannaccone, Laurence R. (1998). “Introduction to the Economics of Religion”. *Journal of Economic Literature* 36(3), 1465–1495.

Inglehart, Ronald, Christian Haerpfer, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Bi Puranen (2014). *World Values Survey: All rounds - Country-pooled datafile version*. Tech. rep. Madrid.

Isphording, Ingo E and Ulf Zölitz (2020). “The Value of a Peer”. *University of Zurich, Department of Economics, Working Paper No. 342*.

Iyer, Sriya (2016). “The New Economics of Religion”. *Journal of Economic Literature* 54(2), 395–441.

Jackson, C. Kirabo (2018). “What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes”. *Journal of Political Economy* 126(5), 2072–2107.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico (2016). “The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms”. *Quarterly Journal of Economics* 131(1), 157–218.

Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel (2020). “School Effects on Socioemotional Development, School-based Arrests, and Educational Attainment”. *American Economic Review: Insights* 2(4), 491–508.

Jackson, C. Kirabo, Cora Wigger, and Heyu Xiong (2021). “Do School Spending Cuts Matter? Evidence from the Great Recession”. *American Economic Journal: Economic Policy* 13(2), 304–35.

Jackson, Margot I. (2009). “Understanding links between adolescent health and educational attainment”. *Demography* 46(4), 671–694.

Jacob, Brian A. (2005). “Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools”. *Journal of Public Economics* 89(5), 761–796.

Jacob, Brian A., Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen (2018). “Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools”. *Journal of Public Economics* 166, 81–97.

Jansen, Philip R, Kyoko Watanabe, Sven Stringer, Nathan Skene, Julien Bryois, Anke R Hammerschlag, Christiaan A de Leeuw, Jeroen S Benjamins, Ana B Muñoz-Manchado, Mats Nagel, et al. (2019). “Genome-wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways”. *Nature Genetics* 51(3), 394–403.

Jayachandran, Seema (2021). “Social Norms as a Barrier to Women’s Employment in Developing Countries”. *IMF Economic Review* 69(3), 576–595.

Jerald, Craig D (2008). “Benchmarking for Success: Ensuring US Students Receive a World-Class Education”. *Washington, DC: National Governors Association*.

Jin, Qiang, Syed Hassan Raza, Muhammad Yousaf, Umer Zaman, and Jenny Marisa Lim Dao Siang (2021). “Can Communication Strategies Combat COVID-19 Vaccine Hesitancy with Trade-Off between Public Service Messages and Public Skepticism? Experimental Evidence from Pakistan”. *Vaccines* 9(7).

Johnson, Carla C (2007). "Effective science teaching, professional development and no child left behind: Barriers, dilemmas, and reality". *Journal of Science Teacher Education* 18(2), 133–136.

Johnson, David R., Christopher P. Scheitle, and Elaine Howard Ecklund (2016). "Conservative Protestantism and Anti-Evolution Curricular Challenges Across States". *Social Science Quarterly* 97(5), 1227–1244.

Johnson, Rucker C. and C. Kirabo Jackson (2019). "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending". *American Economic Journal: Economic Policy* 11(4), 310–49.

Jones, Charles I. (1995). "R & D-Based Models of Economic Growth". *Journal of Political Economy* 103(4), 759–784.

Joravsky, David (1962). "The Lysenko Affair". *Scientific American* 207(5), 41–49.

Jordan, Marty P and Matt Grossmann (2020). "The Correlates of State Policy Project v.2.2." *East Lansing, MI: Institute for Public Policy and Social Research (IPPSR)*.

Jürges, Hendrik, Kerstin Schneider, and Felix Büchel (2005). "The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from Timss Germany". *Journal of the European Economic Association* 3(5), 1134–1155.

Kane, Thomas J, Antoniya M Owens, William H Marinell, Daniel RC Thal, and Douglas O Staiger (2016). "Teaching higher: Educators' perspectives on Common Core implementation". *Cambridge, MA: Center for Education Policy Research*, 9–15.

Keller, Matthew (2014). "Gene × Environment Interaction Studies have not Properly Controlled for Potential Confounders: the Problem and the (Simple) Solution". *Biological Psychiatry* 75(1), 18–24.

Kerr, William R. and William F. Lincoln (2010). "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention". *Journal of Labor Economics* 28(3), 473–508.

Kirkebøen, Lars (2021). "School Value-added and Long-term Student Outcomes". *Statistics Norway Discussion Papers* 970.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad (2016). "Field of Study, Earnings, and Self-Selection". *Quarterly Journal of Economics* 131(3), 1057–1111.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller (2019). "Child Penalties across Countries: Evidence and Explanations". *AEA Papers and Proceedings* 109, 122–26.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007). "Experimental Analysis of Neighborhood Effects". *Econometrica* 75(1), 83–119.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen (2015). "Behavioral economics of education". *Journal of Economic Behavior and Organization* 115, 3–17.

Koellinger, Philipp D. and K. Paige Harden (2018). "Using Nature to Understand Nurture". *Science* 359(6374), 386–387.

Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V. Halldorsson, Gisli Masson, Daniel F. Gudbjartsson, Agnar Helgason, Gyda Bjornsdottir, Unnur Thorsteinsdottir, and Kari Stefansson (2018). "The Nature of Nurture: Effects of Parental Genotypes". *Science* 359(6374), 424–428.

Krueger, Alan B. (1999). "Experimental Estimates of Education Production Functions". *Quarterly Journal of Economics* 114(2), 497–532.

Krueger, Alan B. and Mikael Lindahl (2001). "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39(4), 1101–1136.

Kultusministerkonferenz (2021). *Auswertung Religionsunterricht Schuljahr 2019/20*. Tech. rep. Berlin.

Kweon, Hyeokmoon, Casper A.P. Burik, Richard Karlsson Linnér, Ronald de Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn Paige Harden, Thomas A. DiPrete, and Philipp D. Koellinger (2020). "Genetic Fortune: Winning or Losing Education, Income, and Health". *Tinbergen Institute Discussion Papers* 20-053.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach (2018). "School finance reform and the distribution of student achievement". *American Economic Journal: Applied Economics* 10(2), 1–26.

Laliberte, Jean-William (2021). "Long-term Contextual Effects in Education: Schools and Neighborhoods". *American Economic Journal: Economic Policy* 13(2), 336–377.

Lamarck, Jean-Baptiste Pierre Antoine de Monet de (1838). *Histoire naturelle des animaux sans vertèbres*. Paris: Verdière.

Larson, Edward J. (1999). "The Scopes Trial and the Evolving Concept of Freedom". *Virginia Law Review* 85(3), 503–529.

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

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, 23Andme Research Team, Nguyen-Viet Tuan Anh, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, Mark Alan Fontana, Tushar Kundu, Chanwook Lee, Hui Li, Ruoxi Li, Rebecca Royer, Timshel Pascal N, Raymond K Walters, Emily A Willoughby, Loïc Yengo, Maris Alver, Yanchun Bao, David W Clark, Felix R Day, Nicholas A Furlotte, Peter K Joshi, Kathryn E Kemper, Aaron Kleinman, Claudia Langenberg, Reedik Mägi, Joey W Trampush, Shefali Setia Verma, Yang Wu, Max Lam, Jing Hua Zhao, Zhili Zheng, Jason D Boardman, Harry Campbell, Jeremy Freese, Kathleen Mullan Harris, Caroline Hayward, Pamela Herd, Meena Kumari, Todd Lencz, Jian'an Luan, Anil K Malhotra, Andres Metspalu, Lili Milani, Steven F Lehrer, Philipp D Koellinger, Abedel Abdellaoui, Wouter J Peyrot, Jouke Jan Hottenga, Anna AE Vinkhuyzen, Gonneke Willemsen, Dorret I Boomsma, Cornelia M van Duijn, Brenda Penninx, Danielle Posthuma, and André G Uitterlinden (2018). "Gene Discovery and Polygenic

Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals". *Nature Genetics* 50 (8), 1112–1121.

Lerner, Lawrence S (2000a). "Good and Bad Science in US Schools". *Nature* 407(6802), 287–290.

– (2000b). "Good Science, Bad Science: Teaching Evolution in the States". *Washington, DC: Thomas B. Fordham Foundation*.

Leuven, Edwin and Sturla A Løkken (2020). "Long-term Impacts of Class Size in Compulsory School". *Journal of Human Resources* 55(1), 309–348.

Linden, Sander van der, Anthony Leiserowitz, Geoffrey Feinberg, and Edward Maibach (2015). "The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence". *PloS ONE* 10, e0118489.

Liu, Mengzhen, Yu Jiang, Robbee Wedow, Yue Li, David M Brazel, Fang Chen, Gargi Datta, Jose Davila-Velderrain, Daniel McGuire, Chao Tian, et al. (2019). "Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use". *Nature Genetics* 51(2), 237–244.

Locke, John (1690). *An Essay Concerning Human Understanding*. Hamilton: McMaster University Archive for the History of Economic Thought.

Lott, John R. Jr. (1999). "Public Schooling, Indoctrination, and Totalitarianism". *Journal of Political Economy* 107(6), 127–157.

Lott, Jürgen (2005). "Religionsunterricht in Deutschland". *Jahrbuch für Pädagogik* 2005(1), 143–162.

Loveless, Tom (2014). "The Brown Center Report on Education. How well are American students learning. Part III: A progress report on the Common Core". *The Brookings Institution*.

– (2015). "The Brown Center Report on Education. How well are American students learning? Part II: Measuring effects of the Common Core". *The Brookings Institution*.

– (2016). "The Brown Center Report on Education. How well are American students learning? Part I: Reading and Math in the Common Core Era". *The Brookings Institution*.

– (2021). *Between the State and the Schoolhouse: Understanding the Failure of Common Core.* Harvard Education Press.

Lucas, Robert E. (1988). "On the Mechanics of Economic Development". *Journal of Monetary Economics* 22(1), 3–42.

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

Mankiw, N. Gregory, David Romer, and David N. Weil (1992). "A Contribution to the Empirics of Economic Growth". *Quarterly Journal of Economics* 107(2), 407–437.

Maranto, Robert A. and April Gresham Maranto (2004). "Options for Low-Income Students: Evidence from the States". *Leaving No Child Behind? Options for Kids in Failing Schools*. Ed. by Frederick M. Hess and Chester E. Finn. New York: Palgrave Macmillan US, 63–88.

Marcus, Jan and Vaishali Zambre (2019). “The effect of increasing education efficiency on university enrollment: Evidence from administrative data and an unusual schooling reform in Germany”. *Journal of Human Resources* 54(2), 468–502.

Marks, Jonathan (2012). “Why be Against Darwin? Creationism, Racism, and the Roots of Anthropology”. *American Journal of Physical Anthropology* 149(55), 95–104.

Martin, Alicia R, Christopher R Gignoux, Raymond K Walters, Genevieve L Wojcik, Benjamin M Neale, Simon Gravel, Mark J Daly, Carlos D Bustamante, and Eimear E Kenny (2017). “Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations”. *American Journal of Human Genetics* 100(4), 635–649.

Martinez-Bravo, Monica and Andreas Stegmann (2021). “In Vaccines We Trust? The Effects of the CIA’s Vaccine Ruse on Immunization in Pakistan”. *Journal of the European Economic Association*.

Martschenko, Daphne, Sam Trejo, and Benjamin W. Domingue (2019). “Genetics and Education: Recent Developments in the Context of an Ugly History and an Uncertain Future.” *AERA Open* 5 (1), 1–15.

Matthewes, Sönke Hendrik (2020). “Better Together? Heterogeneous Effects of Tracking on Student Achievement”. *Economic Journal* 131(635), 1269–1307.

Mayr, Ernst (1991). *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought*. Vol. 2. Harvard University Press.

McCleary, Rachel M. and Robert J. Barro (2006a). “Religion and Economy”. *Journal of Economic Perspectives* 20(2), 49–72.

– (2006b). “Religion and Political Economy in an International Panel”. *Journal for the Scientific Study of Religion* 45(2), 149–175.

– (2019). *The Wealth of Religions: The Political Economy of Believing and Belonging*. Princeton University Press.

McCluskey, Neal and Andrew J Coulson (2007). “End It, Don’t Mend It: What to Do with No Child Left Behind. Policy Analysis.” *Cato Institute (Policy Analysis No. 599)*.

Mead, Louise and Anton Mates (2009). “Why Science Standards are Important to a Strong Science Curriculum and How States Measure Up”. *Evolution: Education and Outreach* 2, 359–371.

Meyersson, Erik (2014). “Islamic Rule and the Empowerment of the Poor and Pious”. *Econometrica* 82(1), 229–269.

Mills, Melinda C., Nicola Barban, and Felix C. Tropf (2020). *An Introduction to Statistical Genetic Data Analysis*. Cambridge: MIT Press.

Mincer, Jacob (1958). “Investment in Human Capital and Personal Income Distribution”. *Journal of Political Economy* 66(4), 281–302.

Misco, Thomas (2005). “In response to NCLB: A case for retaining the social studies”. *Essays in Education* 15(1), 2.

Moore, Randy, Murray Jensen, and Jay Hatch (2003a). "The Problems with State Educational Standards." *Science Education Review* 2(3).

– (2003b). "Twenty Questions: What Have the Courts Said about the Teaching of Evolution and Creationism in Public Schools?" *BioScience* 53(8), 766–771.

Muslimova, Dilnoza, Hans van Kippersluis, Cornelius A. Rietveld, Stephanie von Hinke, and S. Fleur W. Meddens (2020). "Dynamic Complementarity in Skill Production: Evidence from Genetic Endowments and Birth Order". *Tinbergen Institute Discussion Paper 2020-082/V*.

National Commission on Excellence in Education (1983). "A Nation at Risk: The Imperative for Educational Reform". *Elementary School Journal* 84(2), 113–130.

National Council for Social Studies (2008). "Curriculum Guidelines for Social Studies Teaching and Learning". *National Council for Social Studies Position Statement*.

National Science and Technology Council (2018). "Charting a Course for Success: America's Strategy for STEM Education". *Report by the Committee on STEM Education of the National Science and Technology Council*.

Neal, Derek and Diane Whitmore Schanzenbach (2010). "Left Behind by Design: Proficiency Counts and Test-Based Accountability". *Review of Economics and Statistics* 92(2), 263–283.

Nicoletti, Cheti and Birgitta Rabe (2014). "School Inputs and Skills: Complementarity and Self-productivity". *IZA Working Paper Series* 8693.

Numbers, Ronald L. (1982). "Creationism in 20th-Century America". *Science* 218(4572), 538–544.

Oates, Wallace E. (1999). "An Essay on Fiscal Federalism". *Journal of Economic Literature* 37(3), 1120–1149.

Opfer, V Darleen, Julia H Kaufman, and Lindsey E Thompson (2016). "Implementation of K–12 state standards for mathematics and English language arts and literacy". *Santa Monica, CA: RAND*.

Oster, Emily (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". *Journal of Business & Economic Statistics* 37(2), 187–204.

Papageorge, Nicholas W and Kevin Thom (2020). "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study". *Journal of the European Economic Association* 18(3), 1351–1399.

Papay, John P. and Matthew A. Kraft (2015). "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement". *Journal of Public Economics* 130, 105–119.

Peri, Giovanni, Kevin Shih, and Chad Sparber (2015). "STEM Workers, H-1B Visas, and Productivity in US Cities". *Journal of Labor Economics* 33(1), 225–255.

Pew Research Center (2009). *Public Praises Science; Scientists Fault Public, Media - Scientific Achievements Less Prominent Than a Decade Ago*.

– (2015). *Public and Scientists' View on Science and Society*.

Pischke, Jorn-Steffen and Till von Wachter (2008). "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation". *Review of Economics and Statistics* 90(3), 592–598.

Plutzer, Eric, Glenn Branch, and Ann Reid (2020). "Teaching Evolution in US Public Schools: A Continuing Challenge". *Evolution: Education and Outreach* 13(1), 1–15.

Polikoff, Morgan S. (2015). "How Well Aligned Are Textbooks to the Common Core Standards in Mathematics?" *American Educational Research Journal* 52(6), 1185–1211.

– (2017). "Is Common Core "Working"? And Where Does Common Core Research Go From Here?" *AERA Open* 3(1), 1–6.

Polleck, Jody N. and Jill V. Jeffery (2017). "Common Core Standards and their Impact on Standardized Test Design: A New York Case Study". *High School Journal* 101(1), 1–26.

Pritchett, Lant and Martina Viarengo (2015). "The State, Socialisation, and Private Schooling: When Will Governments Support Alternative Producers?" *Journal of Development Studies* 51(7), 784–807.

Ramirez, Francisco O. and John Boli (1987). "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization". *Sociology of Education* 60(1), 2.

Reardon, S. F., A. D. Ho, B. R. Shear, E. M. Fahle, D. Kalogrides, H. Jang, and B. Chavez (2021). "Stanford Education Data Archive (Version 4.0)".

Reback, Randall (2008). "Teaching to the rating: School accountability and the distribution of student achievement". *Journal of Public Economics* 92(5), 1394–1415.

Reback, Randall, Jonah Rockoff, and Heather L. Schwartz (2014). "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind". *American Economic Journal: Economic Policy* 6(3), 207–41.

Rios-Avila, Fernando, Pedro H.C. Sant'Anna, Brantly Callaway, and Asjad Naqvi (2021). "csdid and drdid: Doubly robust differences-in-differences with multiple time periods". *Mimeo*.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). "Teachers, Schools, and Academic Achievement". *Econometrica* 73(2), 417–458.

Rivkin, Steven G. and Jeffrey C. Schiman (2015). "Instruction Time, Classroom Quality, and Academic Achievement". *Economic Journal* 125(588), 425–448.

Rockoff, Jonah E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data". *American Economic Review* 94(2), 247–252.

Romer, Paul M. (1990). "Human Capital and Growth: Theory and Evidence". *Carnegie-Rochester Conference Series on Public Policy* 32, 251–286.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm (forthcoming). "Family Disadvantage, Gender and the Returns to Genetic Human Capital". *Scandinavian Journal of Economics*.

Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff (2013). "How Teacher Turnover Harms Student Achievement". *American Educational Research Journal* 50(1), 4–36.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb (2019). “Fast and wild: Bootstrap inference in Stata using boottest”. *Stata Journal* 19(1), 4–60.

Ross, E, N Gerber, H Jarmolowski, K Lakis, N Ledyard, L Staresina, and C Worth (2017). “State teacher policy yearbook”. *Washington, DC: National Council on Teacher Quality*.

Roth, Jonathan and Pedro HC Sant’Anna (2021). “Efficient Estimation for Staggered Rollout Designs”. *Working Paper arXiv:2102.01291*.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio (2013). “Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure”. *American Economic Journal: Economic Policy* 5(2), 251–81.

Rubin, Jared (2014). “Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation”. *Review of Economics and Statistics* 96(2), 270–286.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020). “Integrated Public Use Microdata Series USA: Version 10.0 [dataset]”. *Minneapolis, MN: IPUMS*.

Sacerdote, Bruce (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates”. *Quarterly Journal of Economics* 116(2), 681–704.

– (2014). “Experimental and Quasi-experimental Analysis of Peer Effects: Two Steps Forward?” *Annual Review of Economics* 6(1), 253–272.

Safi, Michael (2018). “Indian education minister dismisses theory of evolution”. *The Guardian*. URL: https://www.theguardian.com/world/2018/jan/23/indian-education-minister-dismisses-theory-of-evolution-satyapal-singh (visited on 07/09/2021).

Savage, Jeanne E. et al. (2018). “Genome-wide Association Meta-analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence”. *Nature Genetics* 50(7), 912–919.

Schmidt, William H. and Richard T. Houang (2012). “Curricular Coherence and the Common Core State Standards for Mathematics”. *Educational Researcher* 41(8), 294–308.

Scholastic (2014). “Teachers’ views on the Common Core State Standards one year later”. *The Bill & Melinda Gates Foundation*.

Schultz, Theodore W. (1961). “Investment in Human Capital”. *American Economic Review* 51(1), 1–17.

Schwerdt, Guido and Ludger Woessmann (2017). “The information value of central school exams”. *Economics of Education Review* 56, 65–79.

– (2020). “Chapter 1 - Empirical methods in the economics of education”. *The Economics of Education (Second Edition)*. Ed. by Steve Bradley and Colin Green. Second Edition. Academic Press, 3–20.

Schwoerbel, Wolfgang (1985). “Das neue Unterrichtsfach Ethik”. *Lehren und Lernen* 2, 9–28.

Selzam, Saskia, Stuart J. Ritchie, Jean-Baptiste Pingault, Chandra A. Reynolds, Paul F. O’Reilly, and Robert Plomin (2019). “Comparing Within- and Between-Family Polygenic Score Prediction”. *American Journal of Human Genetics* 105(2), 351–363.

Settlage, John and Lee Meadows (2002). “Standards-based reform and its unintended consequences: Implications for science education within America’s urban schools”. *Journal of Research in Science Teaching* 39(2), 114–127.

Shakeel, M Danish and Paul E Peterson (2021). “A Half Century of Progress in US Student Achievement: Ethnic and SES Differences; Agency and Flynn Effects”. *Program on Education Policy and Government Report No.: 21-01, Taubman Center for State and Local Government, Harvard Kennedy School*.

Shenhav, Na’ama (2021). “Lowering Standards to Wed? Spouse Quality, Marriage, and Labor Market Responses to the Gender Wage Gap”. *Review of Economics and Statistics* 103 (2), 265–279.

Skoog, Gerald (2005). “The Coverage of Human Evolution in High School Biology Textbooks in the 20th Century and in Current State Science Standards”. *Science & Education* 14(3), 395–422.

Smith-Woolley, Emily, Jean-Baptiste Pingault, Saskia Selzam, Kaili Rimfeld, Eva Krapohl, Sophie von Stumm, Kathryn Asbury, Philip S. Dale, Toby Young, Rebecca Allen, Yulia Kovas, and Robert Plomin (2018). “Differences in Exam Performance between Pupils Attending Selective and Non-selective Schools Mirror the Genetic Differences between Them”. *npj Science of Learning* 3 (1), 1–7.

Solow, Robert M. (1956). “A Contribution to the Theory of Economic Growth”. *Quarterly Journal of Economics* 70(1), 65–94.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie (2019). “Long-term decline in intergenerational mobility in the United States since the 1850s”. *Proceedings of the National Academy of Sciences*.

Sorens, Jason, Fait Muedini, and William P Ruger (2008). “US state and local public policies in 2006: A new database”. *State Politics & Policy Quarterly* 8(3), 309–326.

Sotoudeh, Ramina, Kathleen Mullan Harris, and Dalton Conley (2019). “Effects of the Peer Metagenomic Environment on Smoking Behavior”. *Proceedings of the National Academy of Sciences* 116(33), 16302–16307.

Staff, Toi (2018). “Israeli schools largely avoid teaching evolution”. *Times of Israel*. URL: `https://www.timesofisrael.com/israeli-schools-largely-avoid-teaching-evolution-report/` (visited on 07/09/2021).

Stinebrickner, Ralph and Todd R. Stinebrickner (2014). “A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout”. *Review of Economic Studies* 81(1 (286)), 426–472.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. *Journal of Econometrics* 225(2), 175–199.

Swanson, Christopher B (2005). “Evolution in State Science Education Standards”. *Editorial Projects in Education Research Center*.

Tabellini, Guido (2008). "The Scope of Cooperation: Values and Incentives". *Quarterly Journal of Economics* 123(3), 905–950.

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

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditures". *Journal of Political Economy* 64(5), 416–424.

Trejo, Sam, Daniel Belsky, Jason Boardman, Jeremy Freese, Kathleen Mullan Harris, Pam Herd, Kamil Sicinski, and Benjamin Domingue (2018). "Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health". *Sociological Science* 5, 513–540.

Turkheimer, Eric (2000). "Three Laws of Behavior Genetics and What They Mean". *Current Directions in Psychological Science* 9(5), 160–164.

Van Doorslaer, Eddy and Andrew M Jones (2003). "Inequalities in Self-reported Health: Validation of a New Approach to Measurement". *Journal of Health Economics* 22(1), 61–87.

Ware, Erin B, Lauren L Schmitz, Jessica Faul, Arianna Gard, Colter Mitchell, Jennifer A Smith, Wei Zhao, David Weir, and Sharon LR Kardia (2017). "Heterogeneity in Polygenic Scores for Common Human Traits". *BioRxiv 106062*.

Webber, Ann, Patricia Troppe, Anthony Milanowski, Babette Gutmann, Elizabeth Reisner, and Margaret Goertz (2014). "State Implementation of Reforms Promoted under the Recovery Act. A Report from Charting the Progress of Education Reform: An Evaluation of the Recovery Act's Role". *National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education*.

Wedel, Katharina (2021). "Instruction time and student achievement: The moderating role of teacher qualifications". *Economics of Education Review* 85.

West, Martin and Ludger Woessmann (2010). "'Every Catholic Child in a Catholic School': Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries". *Economic Journal* 120, 229–255.

Winters, Marcus A., Julie R. Trivitt, and Jay P. Greene (2010). "The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam". *Economics of Education Review* 29(1), 138–146.

Wiswall, Matthew and Basit Zafar (2015). "Determinants of College Major Choice: Identification using an Information Experiment". *Review of Economic Studies* 82(2), 791–824.

– (2018). "Preference for the Workplace, Investment in Human Capital, and Gender". *Quarterly Journal of Economics* 133(1), 457–507.

Woessmann, Ludger (2016). "The Importance of School Systems: Evidence from International Differences in Student Achievement". *Journal of Economic Perspectives* 30(3), 3–32.

Wray, Naomi R, Sang Hong Lee, Divya Mehta, Anna AE Vinkhuyzen, Frank Dudbridge, and Christel M Middeldorp (2014). "Research Review: Polygenic Methods and their Application to Psychiatric Traits". *Journal of Child Psychology and Psychiatry* 55(10), 1068–1087.

Xu, Zeyu and Kennan Cepa (2018). “Getting College-Ready during State Transition toward the Common Core State Standards”. *Teachers College Record* 120(6), 1–36.

Yengo, Loic, Julia Sidorenko, Kathryn E Kemper, Zhili Zheng, Andrew R Wood, Michael N Weedon, Timothy M Frayling, Joel Hirschhorn, Jian Yang, Peter M Visscher, et al. (2018). “Meta-analysis of Genome-wide Association Studies for Height and Body Mass Index in 700000 Individuals of European Ancestry”. *Human Molecular Genetics* 27(20), 3641–3649.

Curriculum Vitae

Benjamin W. Arold

09/2017 – 08/2022	Junior Economist and Doctoral Student ifo Institute – Leibniz Institute for Economic Research at the University of Munich
09/2017 – 07/2022	Doctoral Candidate in Economics (Dr. oec. publ.) Munich Graduate School of Economics Ludwig-Maximilians-Universität München
09/2019 – 04/2020	Visiting Research Fellow Program on Education Policy and Government Harvard University
09/2015 – 09/2016	Master of Science in Environmental Economics and Climate Change London School of Economics
09/2014 – 09/2015	Master of Science in Economics University College London
09/2010 – 09/2013	Bachelor of Science in Economics Ludwig-Maximilians-Universität München