

Schmitz, Hendrik; Westphal, Matthias

**Working Paper**

## Early- and later-life stimulation: How retirement shapes the effect of education on old-age cognitive abilities

Ruhr Economic Papers, No. 1146

**Provided in Cooperation with:**

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

*Suggested Citation:* Schmitz, Hendrik; Westphal, Matthias (2025) : Early- and later-life stimulation: How retirement shapes the effect of education on old-age cognitive abilities, Ruhr Economic Papers, No. 1146, ISBN 978-3-96973-329-5, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, <https://doi.org/10.4419/96973329>

This Version is available at:

<https://hdl.handle.net/10419/315483>

**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.*



# RUHR

## ECONOMIC PAPERS

Hendrik Schmitz  
Matthias Westphal

### **Early- and Later-Life Stimulation: How Retirement Shapes The Effect of Education on Old-Age Cognitive Abilities**

## Imprint

---

### Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung

Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics

Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences

Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics

Universitätsstr. 12, 45117 Essen, Germany

### Editors

Prof. Dr. Thomas K. Bauer

RUB, Department of Economics, Empirical Economics

Phone: +49 (0) 234/3 22 83 41, e-mail: [thomas.bauer@rub.de](mailto:thomas.bauer@rub.de)

Prof. Dr. Ludger Linnemann

Technische Universität Dortmund, Department of Business and Economics

Economics – Applied Economics

Phone: +49 (0) 231/755-3102, e-mail: [Ludger.Linnemann@tu-dortmund.de](mailto:Ludger.Linnemann@tu-dortmund.de)

Prof. Dr. Volker Clausen

University of Duisburg-Essen, Department of Economics

International Economics

Phone: +49 (0) 201/1 83 -3655, e-mail: [vclausen@vwl.uni-due.de](mailto:vclausen@vwl.uni-due.de)

Prof. Dr. Ronald Bachmann, Prof. Dr. Almut Balleer, Prof. Dr. Manuel Frondel,

Prof. Dr. Ansgar Wübker

RWI, Phone: +49 (0) 201/81 49-213, e-mail: [presse@rwi-essen.de](mailto:presse@rwi-essen.de)

### Editorial Office

Sabine Weiler

RWI, Phone: +49 (0) 201/81 49-213, e-mail: [sabine.weiler@rwi-essen.de](mailto:sabine.weiler@rwi-essen.de)

### Ruhr Economic Papers #1146

Responsible Editor: Ansgar Wübker

All rights reserved. Essen, Germany, 2025

ISSN 1864-4872 (online) – ISBN 978-3-96973-329-5

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

---

**Ruhr Economic Papers #1146**

Hendrik Schmitz and Matthias Westphal

**Early- and Later-Life Stimulation: How  
Retirement Shapes The Effect of Education on  
Old-Age Cognitive Abilities**

## **Bibliografische Informationen der Deutschen Nationalbibliothek**

---

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

<https://dx.doi.org/10.4419/96973329>

ISSN 1864-4872 (online)

ISBN 978-3-96973-329-5

---

Hendrik Schmitz and Matthias Westphal\*

# Early- and Later-Life Stimulation: How Retirement Shapes The Effect of Education on Old-Age Cognitive Abilities

## Abstract

*We study the interaction of education in adolescence and labor force participation around retirement age and its effect on cognitive abilities for individuals in Europe. Besides a direct long-run effect of education, indirect ones may arise, specifically through labor force participation. We suggest an estimator for causal mediation analysis that accommodates endogeneity and heterogeneous treatment effects and use it to identify indirect effects within the education effect. We find that education raises cognitive abilities by about 8 percent. Among the more educated, labor force participation accounts for 36 percent of the total effect, emphasizing important complementarities between education and labor force participation.*

*JEL-Codes: C31, J14, J24*

*Keywords: Cognitive abilities; causal mediation analysis; marginal treatment effects; education*

*February 2025*

---

\* Hendrik Schmitz, RWI and Paderborn University; Matthias Westphal, RWI and University of Hagen. – We thank Jochem de Bresser, Damon Clark, Martin Huber, Daniel Kühnle, and the participants at the RES conference, the IAAE annual conference, the EEA annual meeting, the Essen Health Conference, the health and population economics seminar at the Leibniz University Hannover, the annual meeting of the German Health Economics Association, and the International Health Economics Workshop Tilburg for helpful comments and discussion. Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged. Detailed acknowledgments for the data at the end of the text. – All correspondence to: Hendrik Schmitz, RWI, Hohenzollernstraße 1–3, 45128 Essen, Germany, e-mail: [hendrik.schmitz@rwi-essen.de](mailto:hendrik.schmitz@rwi-essen.de)

# 1 Introduction

Evidence from neuroscience and economic research makes it appear irrevocable that cognitive abilities decline with age (Kaufman and Horn, 1996; Grady, 2012; Strittmatter et al., 2020). This decline has considerable implications for human interactions, economic choices, and the quality of life per se (see, e.g., Tymula et al., 2013, Christelis et al., 2010, Banks and Oldfield, 2007, Banks et al., 2010, Smith et al., 2010). For instance, dementia—referring to a group of symptoms that originates from the most drastic form of cognitive decline—strongly and increasingly affects many parts of society, ranging from families to the health- and long-term-care systems (Chandra et al., 2023). Additionally, neuroscientific associations between individual lifestyles and cognitive aging suggest that individual behaviors may prevent or at least defer these negative implications (Lindenberger, 2014). Knowledge about the causal determinants of these associations are key for sustainable aging societies.

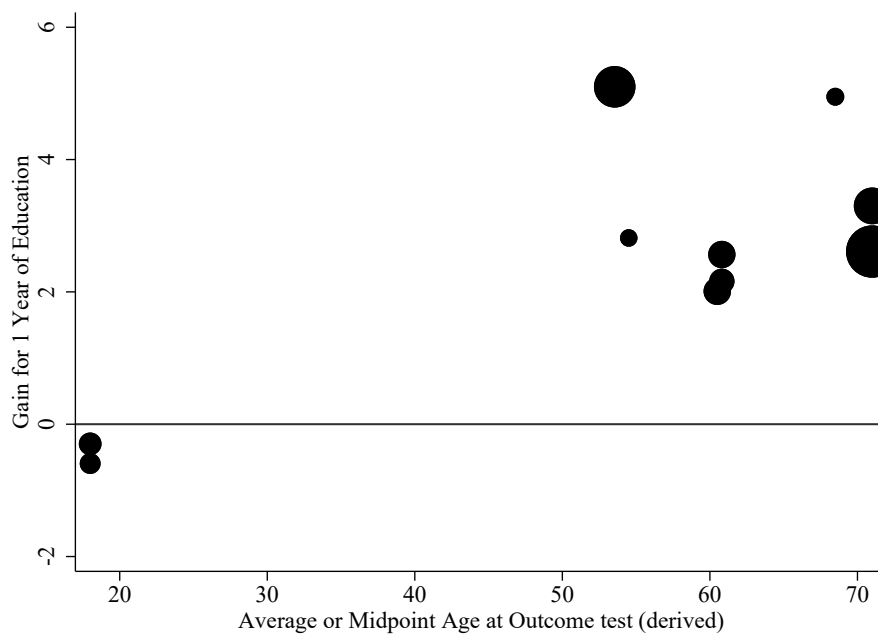
In this paper, we study whether and to which extent the decline and its implications are malleable by education—an important decision in life with many downstream implications over the life course. Although the literature on determinants of skill formation seems to be settled on the fact that hardly anything, including education, impacts *adolescent* cognitive abilities after, say, age 10 (see Heckman, 2007, 2008 for overviews and Cornelissen and Dustmann, 2019 for recent evidence), several studies do find effects of education on the *old-age* cognitive decline (see Figure 1 below). We argue that this seemingly puzzling finding might make sense because of the downstream implications, such as occupational choices, labor supply, and retirement decisions.

To provide some intuition, there are two broad ways in which we think education may affect cognitive decline. First, according to the cognitive reserve hypothesis, individuals with a higher cognitive capacity may be more resilient against an age-related decline (Fratiglioni and Wang, 2007). Thus, if education boosted this capacity, one would expect a slower cognitive decay. Without an effect of education on the level of pre-decline cognitive abilities, however, there would be less scope for this mechanism to work. Second, even if such a direct effect of education on cognitive abilities does not exist, education may still indirectly affect cognitive decline because it may change many other dimensions in life. Take, for instance, labor force participation. As a direct consequence, education influences this decision through occupational choices and the career starting age. While low-educated individuals may be more likely to do routine or physical work, more educated individuals could be more likely to choose more cognitively stimulating jobs. Moreover, if total work experience at retirement was unaffected by education, more educated individuals would need to work until older ages, as they enter the labor market later. Both points would imply a different environment of cognitive stimulation of individuals in their late fifties to sixties, which are typical starting ages for a natural and perceptible cognitive decline.

This second view on the effect of education on cognitive decline is boosted by an intriguing observation that summarizes well the literature on cognitive decline. Using the data from the meta-analysis by Ritchie and Tucker-Drob (2018) and including only results for fluid intelligence (which is a more sensitive leading indicator of a cognitive decline) from studies

in which education is instrumented by some policy change, Figure 1 suggests that the causal effect of education seems to increase with age.<sup>1</sup> Besides these studies on measures of fluid intelligence, [Seblova et al. \(2021\)](#) assess effects on dementia risks directly. They use a sample of very old individuals (with an average age of above 80) and their exposure to a compulsory schooling reform and find that education does not seem to lead to a decreased dementia risk. Thus, there is limited evidence for a direct effect of education (i.e., the cognitive reserve hypothesis does not seem to hold for dementia), but the authors assume that education may alter other risk factors (such as occupational status) in adult life that at least temporarily affect cognitive abilities. We do not claim that this is an exhaustive list of results by age group

Figure 1: The association between age and the effect of education on IQ in compulsory schooling studies



Notes: Based on the meta analysis by [Ritchie and Tucker-Drob \(2018\)](#) complemented by the recent study by [Hampf \(2019\)](#). The remaining studies are [Banks and Mazzonna \(2012\)](#); [Carlsson et al. \(2015\)](#); [Glymour et al. \(2008\)](#); [Gorman et al. \(2017\)](#); [Schneeweis et al. \(2014\)](#). The point size is inversely proportional to the standard error of the respective effect. IQ points are measured by a standardized score with a mean of 100 and a standard deviation of 15.

in the literature, nor that all studies are perfectly comparable with respect to reforms or measures of cognitive abilities. Nevertheless, the effects of education on cognitive abilities in these studies tend to increase in age, particularly in those where some individuals still work. Because individuals with more education are less likely to retire earlier and retirement itself,

<sup>1</sup>Studies that use crystallized intelligence scores (capturing acquired knowledge) as outcomes and, again, compulsory schooling reforms as exogenous variation find a precise positive effect around school-leaving age (see, e.g., [Brinch and Galloway, 2012](#)), but a more ambiguous effect at later ages ([Kamhöfer and Schmitz, 2016](#); [Carlsson et al., 2015](#); [Glymour et al., 2008](#); [Schneeweis et al., 2014](#)). Potentially, this non-amplifying effect along the life course can be explained by a decline in the specific knowledge learned in school.



as is well-documented, has negative effects on cognition<sup>2</sup>, retirement may be the driving force behind Figure 1.

By assessing how retirement empirically alters the effect of education on cognitive decline, we make two main contributions to the literature. First, we develop an estimator for causal mediation analyses that can quantify the role of labor force participation in forming the total effect of education. This flexible estimator accommodates complex forms of endogeneity: endogenous treatment and mediator choices and heterogeneous treatment effects. Conventional IV estimation—including treatment, mediator, and their interaction as regressors, for instance, employed in [Chen et al. \(2019\)](#)—does not identify the contribution of the mediator correctly if there is unobserved heterogeneity and effects for individuals who react to the instrument for the treatment (the compliers) differ substantially from individuals who do not (the always takers and never takers). Our general estimator can be applied to all settings in which the aim is to decompose the causal effect of treatment into a direct and indirect effect that runs through the mediator. Except for [Frölich and Huber \(2017\)](#)<sup>3</sup>, existing methods either address the endogeneity of treatment *or* mediator (as discussed in [Keele et al., 2015](#)) but not of both, or need to make comparably strong assumptions, for instance, to allow one instrument to jointly solve both endogeneity problems ([Dippel et al., 2020](#)).<sup>4</sup>

Basing our estimator on direct instrumental variables estimation, we complement [Frölich and Huber \(2017\)](#) who use a control function approach. We, instead, employ the marginal treatment effect (MTE) framework ([Heckman and Vytlacil, 1999](#)) that explicitly highlights the role of selection in effect heterogeneity. This approach also allows for producing graphics that transparently emphasize the mechanics behind our results while dealing with complex forms of effect heterogeneity. It thereby helps uncover the black box behind estimation approaches to causal mediation analysis.<sup>5</sup> Moreover, while [Frölich and Huber \(2017\)](#) present cases for different combinations of binary/continuous mediator and binary/continuous instrument, they do not address the case that applies here: binary treatment, binary mediator, and two binary instruments. By expanding MTE estimation with binary instruments ([Brinch et al., 2017](#)) to causal mediation analysis, we make this approach applicable also in this (not so uncommon) case. However, if applied to cases with continuous instruments, our approach relies on the same identifying assumptions as [Frölich and Huber \(2017\)](#).

The second main contribution of this paper is one of content. We apply our estimator to precisely identify the extent to which labor force participation moderates the effect of education on cognitive decline, thus analyzing an important pathway of the effect of education. While the literature has explored labor force participation as a potential mechanism of the effect of education, this mechanism has not yet been fully quantified. Typically, by

---

<sup>2</sup>See, for instance, [Rohwedder and Willis \(2010\)](#); [Bonsang et al. \(2012\)](#); [Coe et al. \(2012\)](#); [Celidoni et al. \(2017\)](#); [Mazzonna and Peracchi \(2012, 2017\)](#); [Atalay et al. \(2019\)](#); [Schmitz and Westphal \(2021\)](#).

<sup>3</sup>This approach is also employed by [Salm et al. \(2021\)](#). [Chen et al. \(2020\)](#) adjust this estimator to stochastic frontier models.

<sup>4</sup>Several other recent methodological advances make (sequential) conditional independence assumptions, as documented in [Pearl \(2001\)](#), [Imai et al. \(2010\)](#), [Hong \(2010\)](#), [Tchetgen and Shpitser \(2012\)](#) or [Huber \(2014\)](#).

<sup>5</sup>It also builds the bridge to moderation analyses that explore the interaction of an endogenous treatment with a predetermined covariate. In [Hollenbach et al. \(2024\)](#), for instance, we document that unobserved heterogeneity is also important in this context and employ MTE estimation that considers it.

also reporting the effect of education on labor force participation, the studies report indirect evidence—as we will show, this is only a necessary but by no means sufficient condition of labor force participation being a relevant mediator. Hence, we treat this topic more systematically in this paper and split the (total) effect of education on cognitive abilities into a direct effect and an indirect one going through the channel of labor force participation.

In our analysis, we pool data from SHARE and ELSA on 76,000 observations from several European countries across the years 2002–2017. The data include experimentally collected measures of cognitive abilities (including our main outcome, a word recall test as a proxy for fluid intelligence). We use compulsory schooling reforms and early retirement regulations as sources of exogenous variation and can replicate the effect of education on cognitive abilities as found in the literature. On average, compliers to the compulsory schooling reforms have a larger memory capacity: they recall 7.8 percent more words (24% of an SD) at ages 50–70. Likewise, they are 18 percentage points less likely to be retired on average. We then examine how both effects interact with each other. As a result of our causal mediation analysis, we find about a third of the total effect of education to run through labor force participation. This key finding of our paper may help to resolve the results in previous studies that the short-run effect of education on abilities seems to be small (or absent) while the long-run effect is often estimated to be large. It appears that not education directly but how it affects choices later in life may drive the heterogeneous effects of education on cognitive abilities from prior studies. The additional finding that retirement does not cause a cognitive decline among the less educated compulsory schooling compliers emphasizes an important complementarity between education and old-age labor force participation. Because of these long-run implications of education, education itself may become an even more important target for policy.

This paper proceeds as follows. Section 2 describes the data and institutional regulations before we present the baseline effect of education on cognition in Section 3. In Section 4, we describe the empirical approach how we extend MTE estimation to causal mediation analysis. Estimation results of the mediation analysis are reported in Section 5 while Section 6 concludes.

## 2 Data and institutional set-up

### 2.1 Sample selection and dependent variable

We use data from the Survey of Health Ageing, and Retirement (SHARE) and the English Longitudinal Study of Ageing (ELSA), two large biennial representative micro data sets providing information on health and other socioeconomic characteristics for individuals aged 50 and older.<sup>6</sup> ELSA started in 2002 with 18,000 individuals, while SHARE was initiated in 2004. By now, eight interview waves of SHARE are available, covering information from

---

<sup>6</sup>This Section heavily draws on text from Schiele and Schmitz (2023) and Schmitz and Westphal (2021).

about 140,000 individuals living in Europe.<sup>7</sup> Both data sets are highly harmonized and can be used for pooled analyses.

For our analysis, we use ELSA waves 1–8 and SHARE waves 1, 2, and 4–8 as wave 3 (SHARELIFE) treats different aspects and does not contain the variables of interest.<sup>8</sup> We restrict the sample to individuals between 50 and 70 who are at most eight years above the country-specific early retirement age. In total, we have 80,763 observations from 28,206 individuals living in eight countries.<sup>9</sup>

## Measures of cognitive ability

Cognitive abilities summarize the “ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought” ([American Psychological Association, 1995](#)), where the sum of these abilities is referred to as intelligence. SHARE, HRS, and ELSA offer several potential measures for cognitive abilities: orientation in time, numeracy, verbal fluency, and word recall tests.

In the *word recall test*, the interviewer reads ten words, and the interviewee is asked which words they can remember. The number of words they can recall is counted. This word recall test is done twice: directly after the words are read (immediate recall test) and about 5 minutes later (delayed recall test). The total number of words recalled on these two occasions is added up to yield the word recall test score. This score can range between 0 and 20. Word recall is a measure of episodic memory, which is found to react most strongly to aging ([Rohwedder and Willis, 2010](#)). It is considered a measure of “fluid intelligence”. Broadly speaking, fluid intelligence is the innate cognitive ability, while crystallized intelligence is what people learn in their lifetime (using their fluid intelligence).

In the *verbal fluency test*, respondents are asked to name as many animals as possible in one minute, where the number of animals they can tell becomes their test score. The lower limit is 0, but there is no upper limit (the maximum number in the sample is 100). Verbal fluency measures both fluid and crystallized intelligence, as it is important to know many animals (crystallized knowledge) and remember them quickly (fluid intelligence). Both *recall* and *verbal fluency* only capture specific parts of the multidimensional concept of “cognitive ability”. Our analysis follows much of the recent economic literature and employs *recall* as

---

<sup>7</sup>For comprehensive information on the sampling procedure, questionnaire contents, and fieldwork methodology of HRS, ELSA, and SHARE see [Sonnega et al. \(2014\)](#), [Steptoe et al. \(2003\)](#), and [Börsch-Supan and Jürges \(2005\)](#).

<sup>8</sup>See [Börsch-Supan \(2019a,b,c,d,e,f,g, 2021\)](#); [Brugiavini et al. \(2019\)](#).

<sup>9</sup>These countries are Austria, Germany, Spain, Italy, France, Greece, Czech Republic, and England. Compared to [Schneeweis et al. \(2014\)](#), we include Greece and England. The only country we do not consider from their study is Denmark because there was a coinciding reform of the schooling system (the introduction of a comprehensive schooling system) and because there is some disagreement in the literature about the introduction of the reform (i.e., [Brunello et al., 2009](#) report 1971 whereas [Arendt, 2005](#) states the year 1975). From the other possible countries, we do not include the Netherlands because of doubts about the enforcement of its relatively small 1950 reform (e.g., in [van Kippersluis et al., 2011](#)), Sweden, because school districts could decide to implement the reform before 1969 ([Lundborg et al., 2014](#)), which makes it hard to detect a sudden and clean jump in years of education without any pre-trends, and Belgium, because there also is some disagreement on the timing of the Belgian reform (the source of [Brunello et al., 2016](#) cannot be verified, [Garrouste, 2010](#) does not report this reform. )

our main variable.<sup>10</sup> It has a mean of 10.41 and a standard deviation of 3.43 in our estimation sample.

## 2.2 Education and labor force participation

Our main explanatory variables are binary education measures and the current labor force status. We measure education by years of education and define  $D = 1$  if the number of completed years is at least as large as the compulsory schooling years according to the country-specific rules for the youngest birth cohort in our sample. For example, if compulsory schooling for the birth cohort of 1957 in Austria is nine years, we assign  $D = 1$  to all individuals from Austria with at least nine years of education (irrespective of the birth cohort).  $D$  equals zero if years of education fall below this number. Thus, our binary indicator is a measure of “more” education. Employed as a treatment variable in any IV estimator, this variable enables to condition on individuals affected by compulsory schooling (the compliers). We also report estimates for the total treatment effect where years of education are used as an explanatory variable, and the results are similar. Yet, for our research design, we need a binary treatment.<sup>11</sup>

The upper part of Table 1 shows the distribution of  $D$  in our sample and the average number of years of education for observations with  $D = 0$  and  $D = 1$ . Around three-fourths of individuals are classified as having more education. This is no surprise, as  $D = 0$  only for those individuals in countries with changes in compulsory school years who had the mandatory years of schooling according to the old regime. Individuals with more education, on average, have almost four more years of education. Note, however, that without controlling for the year and country of birth, *inter alia*, this is a crude comparison and does not inform about the magnitude of the first stage.

The lower part of Table 1 informs about the labor force status. We treat individuals as being out of the labor force (that is  $M = 1$ ) if they are either retired, disabled, or not in the labor force due to other reasons. Individuals are in the labor force if they either work part- or full-time, are self-employed, or choose the response option “unemployed” in the respective question. The vast majority of individuals are either employed/self-employed or retired. Note that because we use retirement regulations as an instrument in the subsequent analysis, any effect of  $M$  can be equivalently interpreted as dropping out of the labor force or, more specifically, as effects of retirement for the compliers. We vary the definition of  $M$  in the robustness checks.

Figure 2 shows cognitive ability (left panel) and labor force participation (right panel) by age and treatment status. Both cognitive abilities and labor force participation strongly decline with age. Moreover, there are clearly visible correlations with education. At every age, those with more education have by around one unit higher cognitive abilities. This

<sup>10</sup>See, e.g. Rohwedder and Willis (2010) and Celidoni et al. (2017). Mazzonna and Peracchi (2012) and Coe et al. (2012) use recall and a variety of other measures.

<sup>11</sup>the method could be extended to ordered treatments such as years of education. However, it is unlikely that existing data sets are large enough to identify parameters in such a model with a drastically higher demand for data.

Table 1: Main explanatory variables

	Observations	Mean years of schooling	Realization of $M$
<i>Education (<math>D</math>)</i>			
More education ( $D = 1$ )	57,044	13.00	
Less education ( $D = 0$ )	23,719	9.13	
<i>Labor force status (<math>M</math>)</i>			
Employed/self employed	34,180		0
Unemployed	2,349		0
Retired	34,599		1
Disabled	3,944		1
Not in labor force	5,691		1

Notes: Own calculation based on the pooled selected sample from SHARE and ELSA.

difference seems to increase after the age of 65. In addition, for each age group, the share of individuals out of the labor force is smaller among those with more education.

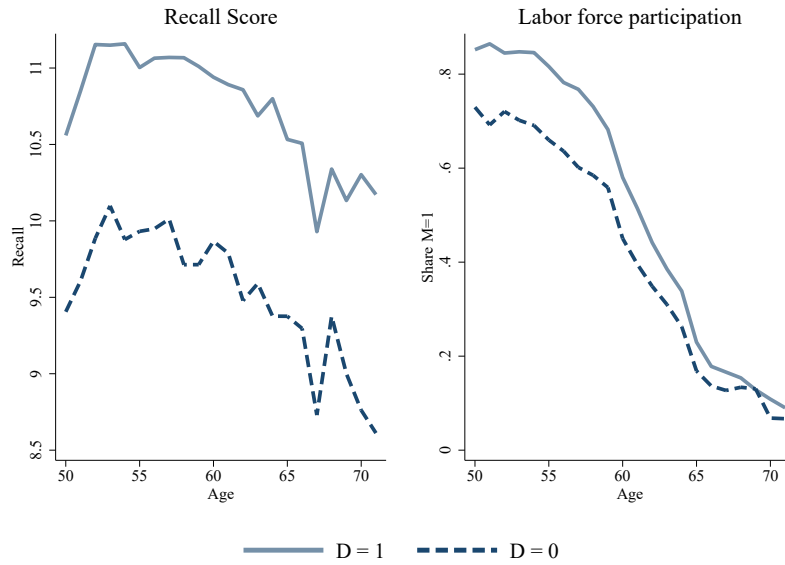


Figure 2: Recall score and labor force participation by age and treatment status

Notes: Own calculation based on the pooled selected sample from SHARE and ELSA. The graph plots unconditional averages by age and treatment status in full years.

## 2.3 Institutional regulations

As is well known, both education and retirement are endogenous when assessing their impact on cognitive ability. We use two established instruments, compulsory schooling reforms, and retirement regulations, to identify the effects of both variables on the outcome. Compulsory schooling reforms, that is, increases in the mandatory years of education, are

not free of critique but are typically considered random from the individual's point of view, at least conditional on certain control variables.

Table 2: Retirement ages and compulsory schooling

	ERA		Compulsory schooling	
	men	women	change in years	pivotal cohort
Austria	60-65	55-60	8-9	1951
Czech Republic	57-60	54-60	8-9	1934
			9-8	1939
			8-9	1947
England	65-66	60-66	10-11	1957
France	60	60	7-8	1923
			8-10	1953
Germany	63	62-63		
BW			8-9	1953
BY			8-9	1955
HB			8-9	1943
HH			8-9	1934
HE			8-9	1953
NI			8-9	1947
NRW			8-9	1953
RLP			8-9	1953
SL			8-9	1949
SH			8-9	1941
Greece	58-60	55-60	6	1963
Italy	57-58	57-58	5-8	1949
Spain	61	61	6-8	1957

Notes: The table shows for each country and gender the Early Retirement Age (ERA) and for each compulsory schooling reform the change in years of compulsory schooling as well as the first cohort affected by the reform. As ERA depends on, e.g., the birth cohort in some countries, we provide the ERA range in our sample for these countries. Information about the compulsory schooling reforms in most countries is taken from [Brunello et al. \(2016\)](#). Additional information about the reforms in Spain, Greece, and England is taken from [Brunello et al. \(2013\)](#). Detailed information on retirement rules for each country is in the supplementary materials.

Table 2 reports, for all countries, years of compulsory schooling before and after a regulatory change and the birth cohort that was first affected (that is, the pivotal cohort). In many cases, these increase from 8 to 9 years, but there is quite some variation. Some countries witnessed more than one reform; specific regions or federal states were subject to different reforms in some countries. We take all of this into account. The pivotal cohorts make clear that some reforms were too early or too late to observe both affected and unaffected individuals within the same country. This is discussed in more detail below. We define the instrument by pooling the information of the compulsory schooling reforms and set  $CS = 1$  for the pivotal cohorts and those born later and  $CS = 0$  for others.

Table 2 also reports early retirement ages (ERA), which are used as instruments for being retired. The early retirement age is when individuals can retire for the first time (as long as



they are not disabled). Early retirement goes along with a penalty on the retirement benefits, which usually is gradually decreased until the official retirement age (ORA) is reached. The institutional rules used to calculate early retirement ages are reported in the supplementary materials. Early retirement ages mainly vary by country, gender, and over time. However, in part they also vary by the individual work history and birth year. There is a slight tendency to increase the early retirement age over time within the countries as a reaction of social policies to the challenges brought along by the demographic change, which we exploit as one source of exogenous variation.

Retirement regulations are a common instrument in the literature, e.g., [Celidoni et al. \(2017\)](#), [Mazzonna and Peracchi \(2012\)](#), [Mazzonna and Peracchi \(2017\)](#). However, in contrast to the previous literature, we only use the *early* but not the *official* retirement age as an instrument and define  $ERA = 1$  if the early retirement age is reached and  $ERA = 0$  if not. Please refer to [Schmitz and Westphal \(2021\)](#), who show that—at least in these data—the first-stage effect of reaching the ORA is not strong enough to use as an instrument once age is properly controlled for. It turns out that ERA seems to be the more important incentive to retire than ORA. We consider the ERA as the onset of a dynamic incentive structure comprising the ORA. We report results of additionally including ORA as an instrument for  $M$  in the robustness checks.

Table 3 shows the interplay of both instruments  $ERA$  and  $CS$ . For identification, we need individuals born before and after the pivotal cohorts and—within both groups—individuals above and below early retirement age. Take, as one example, Austria, the first line of Table 2. We have data for the years 2004–2020 (the ELSA data spans until 2018). In 2016, the pivotal cohort of 1951 turned 65, the (then) ERA for men. Thus, (only) in the most recent wave of data, we observe men from Austria with  $CS = 1$  and  $ERA = 1$ . We do, however, also observe all three other combinations of  $CS$  and  $ERA$ . Since the ERA is 60 in 2016 for women, we have more observations with  $CS = 1$  and  $ERA = 1$  here and, again, also observations with all three other combinations. In the Czech Republic, as another example, compulsory schooling reforms took place so early that we do not observe individuals with  $CS = 1$  and  $ERA = 0$ . Thus, we, in part, need to rely on cross-country variation in institutional settings.

The number of observations for each country and by the treatment  $D$  (education), the mediator  $M$  (being in the labor force), and both instruments are reported in Table A1. Note that we have three countries, the UK, Greece, and Spain, where the respective compulsory schooling reform was too late for individuals in our sample such that no one older than 55 was affected. Those three countries serve as a control group. Having a significant fraction of observations where the instrument is never switched on may help to prevent important drawbacks of difference-in-differences and event-study settings ([Sun and Abraham, 2021](#);

Table 3: When can the pivotal cohorts retire?

Country	Pivotal Cohort	Year surveyed in SHARE/ELSA																		
		Men									Women									
		04	06	08	10	12	14	16	18	20	04	06	08	10	12	14	16	18	20	
Austria	1951	x	x	x	x	x	x	x	o	o	x	x	x	o	o	o	o	o	o	
Czech Republic	1934	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	1939	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	1947	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
England	1933	o	o	o	o	o	o	o	o	—	o	o	o	o	o	o	o	o	—	
	1957	x	x	x	x	x	x	x	x	—	x	x	x	x	x	x	x	o	—	
France	1923	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	1953	x	x	x	x	x	o	o	o	o	x	x	x	x	x	o	o	o	o	
Germany	BW 1953	x	x	x	x	x	x	o	o	o	x	x	x	x	x	x	o	o	o	
	BY 1955	x	x	x	x	x	x	x	o	o	x	x	x	x	x	x	x	o	o	
	HB 1943	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	HH 1934	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	HE 1953	x	x	x	x	x	x	o	o	o	x	x	x	x	x	x	o	o	o	
	NI 1947	x	x	x	o	o	o	o	o	o	x	x	x	o	o	o	o	o	o	
	NRW 1953	x	x	x	x	x	x	o	o	o	x	x	x	x	x	x	o	o	o	
	RLP 1953	x	x	x	x	x	x	o	o	o	x	x	x	x	x	x	o	o	o	
	SL 1949	x	x	x	x	o	o	o	o	o	x	x	x	x	o	o	o	o	o	
	SH 1941	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
Greece	1963	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Italy	1949	x	o	o	o	o	o	o	o	o	x	o	o	o	o	o	o	o	o	
Spain	1957	x	x	x	x	x	x	x	o	o	x	x	x	x	x	x	x	o	o	
		x	stands for: not yet old enough to be eligible to retire																	
		o	stands for: old enough to be eligible to retire																	
		—	stands for: no data																	

Notes: Own illustration based on information in Table 2.

De Chaisemartin and D'Haultfoeuille, 2018; De Chaisemartin and d'Haultfoeuille, 2020).<sup>12</sup>

We test whether our results are affected by these problems in the robustness checks.

<sup>12</sup>These drawbacks could arise in linear models because of an overprediction of the probability to take the treatment (being affected by compulsory schooling) by combining the manifold fixed effects that are often necessary for identification (in our application especially cohort and country fixed effects). Observations at the end of the observation period (the youngest cohorts) in groups that are treated early (countries that introduced compulsory schooling first) are particularly likely to have a predicted first stage value that exceeds one (through the combination of the corresponding fixed effects). De Chaisemartin and d'Haultfoeuille (2020) showed that this overprediction leads to a negative weight for the treatment effect of the corresponding units. Including countries without a compulsory schooling reform in the considered period alleviates these problems.



## 3 Baseline regressions: Effect of education on cognition and retirement

### 3.1 Testing local average treatment effect assumptions

We start with the local average treatment effect of education on cognition and retirement. Before we proceed with the regressions, we carry out some formal IV tests, closely following the procedure in Mourifié and Wan (2017).<sup>13</sup> They suggest a test procedure that jointly tests the hypothesis that an instrumental variable fulfills the exclusion restriction, the independence assumption, and the monotonicity assumption.<sup>14</sup> Comparable tests have been employed by Salm et al. (2021) in their mediation analysis with similar data – they also use cross-country variation with data from SHARE.

We perform the test with  $CS$  as the instrument,  $D$  as the treatment, and cognition/labor force status as outcome variables. Since we only assume independence conditional on control variables, we use the residuals of the regression of the outcome variables on a set of control variables in the test. These control variables – that we use in all regressions of this Section – are: birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects, and a gender dummy. Through birth year-fixed effects and country-fixed effects, we have a difference-in-difference design that enables us to compare the arbitrarily set pivotal compulsory schooling cohorts with marginally older ones. These latter controls for age and wave effects are important to differentiate education (and, later, retirement effects) from the general decline in these age groups. Repetition-fixed effects absorb a potential improvement in the recall score driven by plain familiarization with the test.

Following Mourifié and Wan (2017), we conduct tests for the full sample but also by subgroups. Here, we also test within each country.<sup>15</sup> As a result, shown in columns (1) and (2) of Table 4, the p-values are always larger than 0.20, mostly even 1.00. This holds for both outcome variables and the full sample as well as the single countries. Thus, we do not have evidence to reject the hypothesis that  $CS$  fulfills exclusion restriction, conditional independence, and monotonicity. In anticipation of the next subsection (Figure 3), absent pre-trends in an event-study regression of cognition on the  $CS$  provides additional support for the instrument.

The table also tests the validity of  $ERA$  in the equation of the effect of labor force participation on cognition in column (3). These results will be discussed when we scrutinize the assumptions in the mediation analysis in Section 5.

---

<sup>13</sup>We also adapt their Stata-files, which use the command `clrttest` to our case.

<sup>14</sup>Note that, while Mourifié and Wan (2017) only speak of independence and monotonicity, their test is also triggered when the exclusion restriction is violated.

<sup>15</sup>The latter only works for countries with variation in  $CS$ .

Table 4: IV validity tests

Sample	CS for $D$ in outcome equation (cognition)	CS for $D$ in mediator equation ( $M$ )	ERA for $M$ in outcome equation (cognition)
	(1) p-value	(2) p-value	(3) p-value
All	1.00	1.00	1.00
Austria	1.00	1.00	1.00
Germany	0.29	1.00	1.00
Spain			1.00
Italy	1.00	1.00	1.00
France	1.00	1.00	1.00
Greece			1.00
Czech Republic	0.20	0.98	1.00
England			1.00

Instrument validity tests following [Mourifié and Wan \(2017\)](#). The null hypothesis is that the LATE assumptions hold. Tests are carried out for the full sample and for each country separately. No test carried out in countries without variation in CS.

### 3.2 Effect of education on cognition

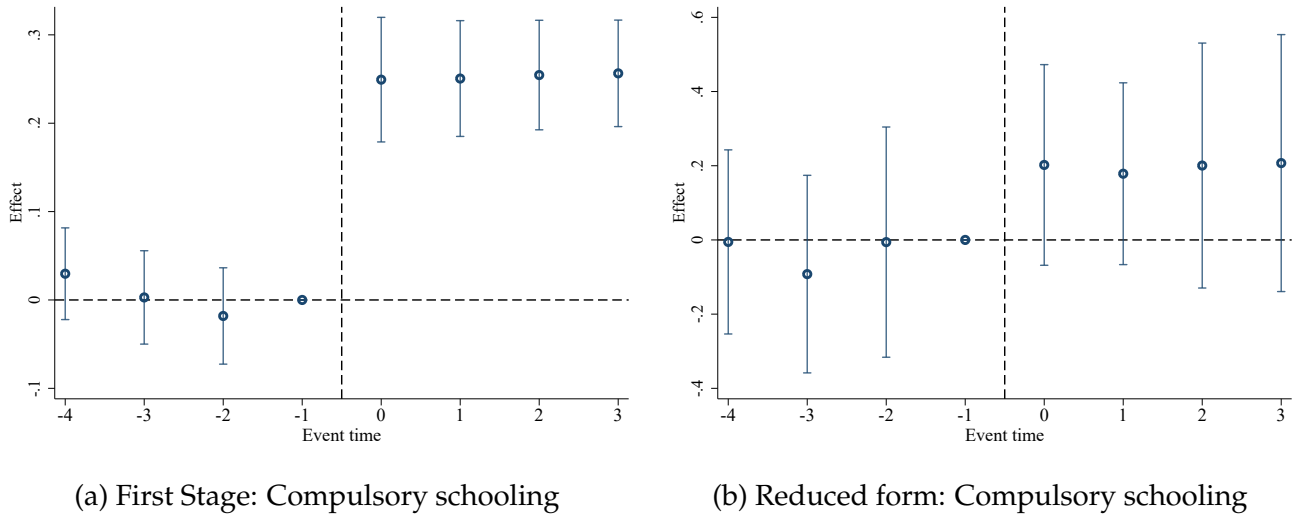
We first present some event-study evidence on the first stage and the reduced form. Both analyses estimate cohort-specific effects relative to the pivotal cohort and its relationship with schooling. The first stage identifies the complier share for the corresponding event years, while the reduced form gives the respective effects on cognitive abilities. Both analyses rely on the following regression, where the event time  $r$  is the normalized birth year ( $r = \text{birth year} - \text{pivotal cohort}$ )

$$Y_{it} = \gamma_{pre} \mathbb{1}[r < -4] + \sum_{\substack{j=-4 \\ j \neq -1}}^{\infty} \gamma_j \mathbb{1}[r = j] + \mathbf{X}'_{it} \boldsymbol{\beta} + \varepsilon_{it},$$

where the vector  $\mathbf{X}_{it}$  consists of the controls mentioned in the previous subsection. For the first stage, we use the education indicator  $D_{it}$  as the dependent variable, whereas the reduced form uses cognitive abilities  $Y_{it}$ . The coefficients of interest are  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$ , which measure the effects of compulsory schooling for the different cohorts relative to  $r = -1$  (our reference category). Additionally, we are interested in the pre-event coefficients  $\gamma_{-4}$ ,  $\gamma_{-3}$ , and  $\gamma_{-2}$  as these detect any deviation from a common trend in the outcome and treatment before the introduction of the reform. We plot these coefficients in Figure 3. We do not plot coefficients for higher event times to ensure that a homogeneous sample of countries contributes to the effects. Nonetheless, we include saturated event-time effects for all periods larger than 3 (up to  $r = 22$ ) to avoid contamination from other periods. The same holds for event times less than -4. For these, we include a joint indicator.

In the left panel of Figure 3, we present the first stage, which is estimated precisely. There are no pre-trends and a pronounced and persistent upward shift in education when a compulsory schooling reform took effect. The jump of the effects between  $r = -1$  and  $r = 0$  means that, on average, 24 percent of all individuals had to extend their schooling because

Figure 3: Compulsory schooling: first stage and reduced form on cognitive abilities



Notes: Event time is birth cohort minus pivotal cohort. The vertical lines around the dots indicate 95 percent confidence intervals. Standard errors clustered on the birth year-country level.

of the compulsory schooling reforms.<sup>16</sup> This is the group of compliers in the IV terminology. As never-takers cannot exist in this setting, the remaining 76 percent are always-takers.<sup>17</sup> Although the compliers are a rather small group of individuals, it is a potentially very interesting one: compliers, in our case, only take the minimum necessary level of education (most likely also with the reform). Thus, we will identify effects for individuals with a low preference for education.

Figure 3b shows the effects on cognitive abilities. In the four years before the reform, no differential effects are visible, suggesting our estimates later do not capture some specific cohort trends. Starting with the first affected compulsory-schooling cohort, we see an elevated level of cognitive skills, which stays around 0.2 (amount of words recalled more) for the subsequent periods. Although neither post-compulsory-schooling effect is significant at the 95% significance level when taken on its own, a joint estimation reveals a precise reduced-form effect, presented below. In total, the event-study evidence suggests that the compulsory schooling reforms sharply increased years of schooling for the affected cohorts without any detectable pre-trends. Thus, these changes most likely explain similar old-age cognitive ability trajectories for cohorts around the introduction of the reforms.

Table 5 reports regression results of both OLS and 2SLS (including the first stage) of cognitive abilities on education and controls. OLS in column (1) repeats the finding from Figure 2 that already showed higher cognitive abilities for those with more education. On average, those with more education score almost 1.4 points higher on the cognitive abilities test. The first stage regression of  $D$  on the instrument  $Z_1$  in column (2) aggregates the numbers already seen in Figure 3a to 0.24, i.e., 24 percent of all individuals are compliers because they need to adjust their otherwise preferred amount of schooling due to the

<sup>16</sup>The exact number of 24 percent follows from Table 5 below.

<sup>17</sup>We adjusted the schooling years to compulsory years in the few cases where the self-reported years are lower.

compulsory schooling reform. We can contrast this finding with the reduced-form result in column (4). Compulsory schooling raises the number of recalled words in the overall population by 0.2. This effect is precisely estimated and significant at the 5 percent level. Finally, the effect of more education on cognitive abilities in column (3) is around 0.8. The compliers to the compulsory schooling reforms recall 0.8 more words due to more education. This is a considerable amount. To interpret this effect size, we could, for instance, relate it to the level or the trend plotted in Figure 2. The individuals with  $D = 0$  are a more appropriate comparison, as these are all potential compliers with the same level of schooling that the treated compliers would have preferred. Compulsory schooling could raise cognitive abilities around age 55-60 by about 8 percent or by 24 percent of a standard deviation in the full sample. Another way of interpretation is the comparison to the general age-related decline, which is about one word in 15 years; see Figure 2. Thus, compulsory schooling could roll back the age-related decline in recall at age 70 by approximately 12 years. In Table A2 in the Appendix, we add regression results with a different treatment variable, namely years of education. The results are comparable.

Table 5: Regression results: effects of education on cognition

	OLS (1)	First stage (2)	2SLS (3)	Reduced Form (4)
More education ( $D$ )	1.423*** (0.052)		0.811*** (0.306)	
Post CS-reform ( $Z_1$ )		0.244*** (0.020)		0.198** (0.081)
Control variables	yes	yes	yes	yes

Number of observations in each regression: 80,763. Additional control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects and male. Standard errors in parentheses clustered on birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 3.3 Potential mechanism: Effect of education on retirement

We now estimate the effect of education on retirement as a potential channel for the effect of education on cognitive abilities. As Table 6 shows, there is indeed a considerable effect. The compliers in our sample are 17.7 percentage points less likely to be retired due to compulsory schooling. This effect is even larger than the OLS difference, demonstrating that individuals at the lowest educational margins adjust their labor supply in response to more education.

Moreover, the effect of retirement on cognition is well-documented, see, e.g. Schmitz and Westphal (2021). Their study also shows that there are no pre-trends in cognitive abilities or the probability of being in the labor force before retirement eligibility.

Table 6: Regression results: Effect of  $D$  on  $M$ 

	OLS (1)	2SLS (2)
More education ( $D$ )	−0.082*** (0.008)	−0.177*** (0.050)
Control variables	yes	yes

Number of observations in each regression: 80,763. Additional control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects and male. Standard errors in parentheses clustered on birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## 4 Empirical approach to mediation analysis

We have seen that education affects cognitive ability at older ages. We have also seen that education affects labor force participation. This raises the important question of how much of the overall effect can be explained by the labor force participation channel. In this section, we present an empirical approach that we will use to answer this question. In the next section, we apply this approach and estimate that labor force participation can account for around one-third of the effect.

### 4.1 Notation and parameters of interest

Our empirical approach starts from the traditional potential outcome model, where  $Y^1$  and  $Y^0$  are the potential outcomes—cognitive abilities—with and without treatment. The binary treatment  $D$  is “more” education. The observed outcome  $Y$  either equals  $Y^1$  if an individual received treatment or  $Y^0$  in the absence of the treatment. We suppress the individual identifier  $i$ . We want to find out what drives the effect of  $D$  on the outcome variable  $Y$ . Technically, we want to decompose the effect of education on cognition,  $E(Y^1 - Y^0)$  into components caused by potential mediating factors and a remaining part that may be the direct effect of education.

In this paper, we focus on one mediating factor—the binary mediator variable  $M$ —but the setting can potentially be extended to accommodate any number of those factors. This mediator—retirement—may affect  $Y$  but may also be affected by  $D$ . It, likewise, is the realization of one of the two potential outcomes  $M^1$  and  $M^0$ , where  $M^1$  is labor force status with  $D = 1$  and  $M^0$  is labor force status with  $D = 0$ . The potential outcome then reads  $Y^{jk}$  where  $j \in \{0, 1\}$  denotes the educational choice  $D$  and  $k \in \{0, 1\}$  denotes the retirement choice  $M$ . For instance,  $Y^{11}$  is the potential outcome if the treatment and the mediator are chosen. In the mediation literature, researchers typically aim not to keep  $M$  fixed in the potential outcomes at a certain value but allow  $M$  to vary as a response to  $D$ . Thus, the usual potential outcomes of interest are  $Y^{jM^k}$ .

In this notation, the treatment effect of  $D$  on  $Y$  that leaves implicit the channel  $M$  is called the total treatment effect (TTE) and is defined as follows (Frölich and Huber, 2017):

$$\text{TTE} = E(Y^1 - Y^0) = E(Y^{1M^1} - Y^{0M^0})$$

In Section 3, we estimate a local average total treatment effect of the effect of  $D$  on  $Y$ . The total treatment effect can be disaggregated into two components: the direct effect of the treatment on the outcome and the indirect effect that goes through the mediator. The direct treatment effect (DTE) holds the mediator constant at either  $M^0$  or  $M^1$  and is therefore defined as

$$\text{DTE}(k) = E(Y^{1M^k} - Y^{0M^k})$$

where  $k \in \{0, 1\}$ . Thus, it closes down the effect  $D$  has on  $Y$  via  $M$ . The indirect treatment effect (ITE) is defined vice versa. It fixes the treatment at  $D = j$  and only varies the mediator from  $M^0$  to  $M^1$ :

$$\text{ITE}(j) = E(Y^{jM^1} - Y^{jM^0})$$

Adding and subtracting  $Y^{0M^1}$  from the TTE yields

$$\begin{aligned} \text{TTE} &= E(Y^{1M^1} - Y^{0M^0}) \\ &= E(Y^{1M^1} - Y^{0M^1} + Y^{0M^1} - Y^{0M^0}) \\ &= E(Y^{1M^1} - Y^{0M^1}) + E(Y^{0M^1} - Y^{0M^0}) = \text{DTE}(1) + \text{ITE}(0) \end{aligned} \quad (1)$$

Analogously, adding and subtracting  $Y^{1M^0}$  yields  $\text{TTE} = E(Y^{1M^0} - Y^{0M^0}) + E(Y^{1M^1} - Y^{1M^0}) = \text{DTE}(0) + \text{ITE}(1)$ . Differences between TTE and DTE(0) or DTE(1) arise only if two conditions hold jointly. First,  $D$  needs to change the individuals' choice behavior of  $M$ , i.e.,  $M^1 \neq M^0$  for a relevant subgroup of individuals. Second, these changes in the choice behavior must induce meaningful changes in the  $k$ -part of the potential outcomes, i.e.,  $Y^{jM^1} - Y^{jM^0} \neq 0$  for the same group of individuals.

## 4.2 Simple estimation approaches, their problems, and a sketch of a solution

To calculate the treatment parameters, we need to estimate the four magnitudes  $E(Y^{1M^1})$ ,  $E(Y^{1M^0})$ ,  $E(Y^{0M^1})$ , and  $E(Y^{0M^0})$ . Take, as an example,  $E(Y^{1M^0})$ . It is an inherently hypothetical outcome, but on average, we have  $E(Y^{1M^0}) = \Pr(M^0 = 1) \cdot E(Y^{11}) + \Pr(M^0 = 0) \cdot E(Y^{10})$ . Estimating  $\Pr(M^0 = 1)$ ,  $\Pr(M^0 = 0)$ ,  $E(Y^{11})$  and  $E(Y^{10})$  is straightforward if we have two independent random assignments of  $D$  and  $M$  and full compliance. Assuming a linear model and abstracting from further control variables, we can run the following OLS regressions and use the estimated parameters to calculate all treatment effects as follows:

$$Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \epsilon \quad (2)$$

$$M = \gamma_0 + \gamma_D D + u \quad (3)$$

Then, for example,  $E(Y^{11}) = \delta_0 + \delta_D + \delta_M + \delta_{DM}$  or  $Pr(M^0 = 1) = \gamma_0$ . While these are two examples for illustration, the two regressions allow us to calculate all numbers needed to estimate TTE, ITE, and DTE.

Things become more problematic when  $D$  and  $M$  are choice variables and, thus, endogenous. In principle, instruments for  $D$  and  $M$  could be used and separate 2SLS regressions could be estimated. [Nicoletti et al. \(2023\)](#) address endogeneity problems in this spirit in their causal mediation analysis of the effects of maternal labor force participation on child abilities. However, if treatment effects are heterogeneous, two different complier groups (with respect to the different instruments for  $D$  and  $M$ ) identify the parameters, complicating the interpretation of the treatment effects. This interpretation may even be confounded when  $M$  is a predetermined covariate, as documented by [Hollenbach et al. \(2024\)](#). Denote  $Z_D$  the instrument for  $D$ ,  $Z_M$  the instrument for  $M$ ,  $C_D$  the group of compliers to  $Z_D$  and  $C_M$  the group of compliers to  $Z_M$ . With separate 2SLS, we get estimates of, say,  $E(Y^{11}|C_D, C_M)$  that is, of some subgroup of compliers to the  $D$ -instrument and the  $M$ -instrument and  $Pr(M^0 = 1|C_D)$ . If the complier groups differ and have different effects, the resulting ITE and DTE are no well-defined parameters for a clear population.

We now sketch our proposed solution and how we aim to estimate our parameters of interest: TTE, DTE, and ITE for the compliers to the  $D$  instrument. Thus, we aim at estimating  $E(Y^{1M^1}|C_D)$ ,  $E(Y^{1M^0}|C_D)$ ,  $E(Y^{0M^1}|C_D)$ , and  $E(Y^{0M^0}|C_D)$ . Details will be discussed in the next subchapter. To begin, take a step back and ignore mediation analysis for a moment. Here we know:

- If  $D$  is exogenous, we can simply estimate  $E(Y^1) = E(Y|D = 1)$  and  $E(Y^0) = E(Y|D = 0)$ .
- If  $D$  is endogenous but we have an instrument, [Imbens and Rubin \(1997\)](#) show how we can estimate potential outcomes for compliers, that is  $E(Y^1|C_D)$  and  $E(Y^0|C_D)$ .
- [Carneiro and Lee \(2009\)](#) go a step further and show how one can estimate the distribution of potential outcomes for the full population normalized to the unit interval, that is  $E(Y^1|U_D = u_D)$ ,  $E(Y^0|U_D = u_D)$  where  $u_D \in [0, 1]$  denotes the cumulative share of individuals up to a certain rank. More recently,  $E(Y^1|U_D = u_D)$ ,  $E(Y^0|U_D = u_D)$  have been termed marginal treatment response functions (MTR, see [Mogstad et al., 2018](#)). In a second step, one can aggregate up all  $E(Y^1|U_D = u_D)$  and  $E(Y^0|U_D = u_D)$  to derive  $E(Y^1)$  and  $E(Y^0)$  for the full population (not just for the compliers).

Now, turn back to mediation analysis, but assume, again, that  $D$  is exogenous but  $M$  is endogenous:

- We can use subsample  $D = 1$  and instrument  $M$  with  $Z_M$  to estimate  $E(Y^{11}|U_M = u_M)$  and  $E(Y^{10}|U_M = u_M)$  and to derive  $E(Y^{11})$  and  $E(Y^{10})$ .
- We can use subsample  $D = 0$  and instrument  $M$  with  $Z_M$  to estimate  $E(Y^{01}|U_M = u_M)$  and  $E(Y^{00}|U_M = u_M)$  and to derive  $E(Y^{01})$  and  $E(Y^{00})$ .

Finally, since  $D$  is not random, we solve this using instrument  $Z_D$  and adapt [Imbens and Rubin \(1997\)](#) to get

- $E(Y^{11}|C), E(Y^{10}|C)$
- $E(Y^{01}|C), E(Y^{00}|C)$

Together with estimates of  $Pr(M^0 = 1|C_D)$  and  $Pr(M^1 = 1|C_D)$ , we can derive  $E(Y^{1M^1}|C_D)$ ,  $E(Y^{1M^0}|C_D)$ ,  $E(Y^{0M^1}|C_D)$ , and  $E(Y^{0M^0}|C_D)$  – and, thus TTE, DTE and ITE for the compliers to  $Z_D$

### 4.3 More formal derivation, identifying assumptions, and estimation procedure

We derive this more formally now and use a generalized Roy model in which individuals weigh their costs and benefits and choose  $M^j = 1$  if the net benefit is positive (e.g., see [Heckman, 2010](#)). Recall that choosing  $M$  may depend on the initial assignment of  $D$ . Let  $Y^{j1} = \mu^{j1} + U^{j1}$  and  $Y^{j0} = \mu^{j0} + U^{j0}$  be the potential outcome functions that can be separated into functions of observables ( $\mu^{j1}$  and  $\mu^{j0}$ ) and unobservables ( $U^{j1}$  and  $U^{j0}$ ). The choice of  $M$  induces costs that can, likewise, be separated into an observable and an unobservable part:  $C^j(Z_M) = \mu_C^j(Z_M) + v^j$ . The instrument  $Z_M$  does not affect the potential outcomes but the cost function and, via this, the choice of  $M$ . Individuals choose  $M = 1$  if expected benefits outweigh expected costs:

$$\begin{aligned}
M^j &= \mathbb{1} \left[ Y^{j1} - Y^{j0} \geq C^j(Z_M) \right] \\
&= \mathbb{1} \left[ \mu^{j1} - \mu^{j0} - \mu_C^j(Z_M) \geq U^{j1} - U^{j0} - v^j \right] \\
&= \mathbb{1} \left[ \mu_M^j(Z_M) \geq \mathcal{V}^j \right] \\
&= \mathbb{1} \left[ F_{\mathcal{V}} \left( \mu_M^j(Z_M) \right) \geq F_{\mathcal{V}} \left( \mathcal{V}^j \right) \right] \\
&= \mathbb{1} \left[ \Pr(M = 1|Z_M) \geq U_M^j \right] \quad \forall j \in \{0, 1\}.
\end{aligned} \tag{4}$$

The second step collects observed components on the left and unobserved components on the right side of the inequality, respectively. Subsequently, the inequality is simplified by renaming  $\mu_M^j(Z_M) = \mu^{j1} - \mu^{j0} - \mu_C^j(Z_M)$  for the observed and  $\mathcal{V}^j = U^{j1} - U^{j0} - v^j$  for the unobserved components. Finally, we monotonically transform both sides of the inequality by  $F_{\mathcal{V}}(\cdot)$ . This transformation yields the ranks ( $U_M^j$ ) when applied to  $\mathcal{V}$  and the probability of retirement when applied to the index of observed factors ( $\mu_M^j$ ). This probability is referred to as the propensity score for any binary treatment. By definition,  $U_M^j$  ranges between 0 and 1 and represents the respective threshold value for an individual with a given set of observable factors (that determine the propensity score) to select into the treatment, here  $M = 1$ . At a specific threshold  $p \in (0, 1)$ , where  $U_M^j = p$ , individuals are indifferent between taking the treatment or not and reveal their taste or distaste for  $M$  by opting for it at a specific value of the propensity score. Applying these principles to individuals with a propensity score close to zero, we can infer their value of  $U_M^j$  for individuals who nonetheless take  $M = 1$ . It has to



be even smaller than the propensity score for everyone, and hence  $U_M^j$  measures an inverse score of the revealed preferences.<sup>18</sup>

Note that the last step of Eq. (4) shows how to estimate the two quantities  $E(M^j|U_M^j)$ . For this, we just need to determine the fraction of estimated propensity score values larger than the respective  $U_M^j$  value at which we evaluate:  $E(\mathbb{1}[\Pr(M = 1|Z_M) \geq U_M^j]|U_M^j)$ . By definition,  $E(M^j|U_M^j)$  has to decay along  $U_M^j$ . In the MTE literature, these quantities are known as the weights for aggregating the MTEs to a single ATT parameter. For our mediation analysis, these curves are weights for aggregating the  $Y^{j1}$  and  $Y^{j0}$  quantities into the relevant  $Y^{jM^0}$  and  $Y^{jM^1}$  outcomes.

The index of the unobserved factors plays a crucial role in gaining more insights into how treatment selection is correlated with the effects it evokes. We define

$$\text{MTR}^{jk}(p) = E(Y^{jk} | U_M = p)$$

Below, we show how we estimate this and how instruments are key for its identification.

Notice that although the mediating state generally may depend on the treatment state ( $k = k(j)$ ), conditional on  $p$ , marginal variations in the treatment state  $\partial E(Y^{jk} | U_M = p) / \partial j$  do not induce changes in the mediator state. This is because conditional on  $p$ , individuals are indifferent about the mediator so that  $\partial k(j) / \partial j = 0$ .

### Estimation of the MTRs and identifying assumptions

We believe that it is instructive to, again, start with a completely random and full-compliance  $D$  when we discuss the estimation of  $\text{MTR}^{jk}(p)$  before we estimate the MTRs for compliers to  $Z_D$  which we will term  $\text{MTR}_C^{jk}(p)$ .

*Assumption 1:* We have an instrument  $Z_M$  that fulfills the typical LATE assumptions: strength with respect to  $M$ , conditional independence of the potential outcomes  $Y^{jk}$ , exclusion restriction, and monotonicity (“no defiers”).

Transferred to our setting, [Carneiro and Lee \(2009\)](#) and [Brinch et al. \(2017\)](#) (in their “separate estimation approach” of the marginal treatment effect) show that with such an instrument, we can express the MTR at a specific value of the propensity score  $P(Z_M) = \Pr(M = 1|Z_M, X) = p$  as

$$\begin{aligned} \text{MTR}^{11}(X = x, U_M = p) &= E(Y|X = x, P(Z_M) = p, D = 1, M = 1) \\ &\quad + p \frac{\partial E(Y|X = x, P(Z_M) = p, D = 1, M = 1)}{\partial p} \end{aligned} \quad (5)$$

$$\begin{aligned} \text{MTR}^{10}(X = x, U_M = p) &= E(Y|X = x, P(Z_M) = p, D = 1, M = 0) \\ &\quad - (1 - p) \frac{\partial E(Y|X = x, P(Z_M) = p, D = 1, M = 0)}{\partial p} \end{aligned} \quad (6)$$

<sup>18</sup>Also note that our approach can allow  $U_M^j$  to differ by treatment state:  $U_M^1 \neq U_M^0$ . Hence, we do not need to impose any rank invariance restriction for the mediation analysis. Of course, education may change the preference ordering for retirement. We document this in the simulation exercise of Section B in the Appendix, where we assume  $U_M^1$  and  $U_M^0$  to be uncorrelated, thereby ensuring that ranks may differ by treatment state.

Note that while we left control variables implicit before, we now make this explicit by conditioning on the vector of observables  $\mathbf{X}$ . These  $\mathbf{X}$  are used in the propensity score estimation, but, for simplicity, we write  $P(Z_M)$  only. The expectation of  $Y$ , conditional on  $\mathbf{X}$ ,  $p$ , and the mediator status  $M = 1$  can be expressed as

$$E(Y|\mathbf{X} = \mathbf{x}, P(Z_M) = p, M = 1, D = 1) = \mu^{11}(\mathbf{x}) + f^{11}(p), \quad (7)$$

where  $f^{11}(p)$  is an unknown function of  $p$ . Likewise for  $M = 0$

$$E(Y|\mathbf{X} = \mathbf{x}, P(Z_M) = p, M = 0, D = 1) = \mu^{10}(\mathbf{x}) + f^{10}(p). \quad (8)$$

Estimation of the MTR is, then, straightforward. We stratify the sample by  $D$  and  $M$ , specifying  $\mu^{jk}(\mathbf{X})$  as a linear function of  $\mathbf{X}$  plus linear interactions with  $p$  captured by the parameter vectors  $\gamma^{jk}$  and  $\delta^{jk}$ , respectively. Hence, we estimate the parameters in the regression  $Y = \mathbf{X}'\gamma^{jk} + \mathbf{X}'p\delta^{jk} + f^{jk}(p) + \varepsilon$  and calculate the MTR according to Eq. (5) and Eq. (6). Estimation could be semi-parametric as, e.g., in [Westphal et al. \(2022\)](#) where no functional form of  $f^{jk}(p)$  is assumed and the estimator suggested by [Robinson \(1988\)](#) is employed. Alternatively, the shape of  $f^{jk}(p)$  is approximated by including polynomials of  $p$ . The  $\text{MTR}^{0k}$  are estimated equivalently using the subsample of  $D = 0$ .

Since we have a binary instrument, the propensity score will only take on two values, conditional on  $\mathbf{X}$ . This means that we also need variation in  $\mathbf{X}$  to receive a continuous  $p$  and to identify the MTR without making the strong assumption that the MTR is linear, see [Brinch et al. \(2017\)](#). This can be achieved by an additive separability assumption. While this assumption is already included in our specification of potential outcomes above, we spell it out here for completeness.

*Assumption 2:* Additive separability between observed and unobserved heterogeneity in the MTR. This means that  $E(Y^{jk}|U_M, \mathbf{X} = \mathbf{x}) = \mu^{jk}(\mathbf{x}) + E(U^{jk}|U_M)$ . Note that this assumption is not an indispensable precondition for our method in general; we can also allow for non-parametric interactions between  $\mathbf{x}$  and  $E(U^{jk})$  as [Frölich and Huber \(2017\)](#). It only needs to hold once we have a binary instrument.

This assumption restricts the slope (but not the level) of the MTR to be equal across different  $\mathbf{X}$  cells, i.e., treatment effects can vary with  $\mathbf{X}$  and  $U_M$ , but not along their interaction ([Brinch et al., 2017](#)). With this assumption, we can additionally exploit variation in  $\mathbf{X}$  for the full support condition and identify local effects at different values of  $P(Z_M)$  for compliers to  $Z_M$ . This assumption is standard in applied work. For instance, the conventional two-stage least squares approach that does not interact with the treatment  $D$  with  $\mathbf{X}$  implicitly makes an even stronger assumption (that treatment effects are additively separable in  $D$  and  $\mathbf{X}$ ) without further discussion. Moreover, MTEs are generally derived and estimated with this implicit separability assumption. Without it, one would need to separately estimate MTEs for all possible combinations of  $\mathbf{X}$ , which quickly becomes infeasible. Examples of MTE applications exploiting additive separability include, for instance, [Carneiro et al. \(2011\)](#); [Brinch et al. \(2017\)](#); [Nyblom \(2017\)](#); [Cornelissen et al. \(2018\)](#). Note that there are possibly other ways to estimate non-linear MTRs with binary instruments. [Mogstad et al. \(2018\)](#) derive

MTR- and MTE-bounds using flexible Bernstein-polynomials as functional forms of MTRs and MTEs. This has also been applied by [Rose and Shem-Tov \(2021\)](#) and [Hollenbach et al. \(2024\)](#) in contexts other than mediation analysis.

Next, as our treatment, education is endogenous, we allow self-selection into  $D$  and adapt the approach of [Imbens and Rubin \(1997\)](#) to our setting.

*Assumption 3:* We have an instrument  $Z_D$  that fulfills the typical LATE assumptions: strength with respect to  $D$ , conditional independence of the potential outcomes  $Y^{jk}$ , exclusion restriction, and monotonicity (“no defiers”).

For pedagogic reasons, we first show how to estimate the total treatment effect, that is  $E(Y^1 - Y^0|C_D)$ , leaving the potential channel  $M$  implicit.<sup>19</sup> We start by estimating the share of  $Z_D$ -compliers by the following first-stage regression of  $D$  on the instrument  $Z_D$  and a set of control variables, possibly to justify the conditional independence assumption of the instrument. Throughout this exposition, all control variables enter as mean-centered values so that  $E(X) = 0$  (so that the intercept of all regressions gives the respective outcome mean when the covariates are fixed at the mean).

$$D = \pi_0 + \pi_1 Z_D + \mathbf{X}'\boldsymbol{\pi}^X + u \quad (9)$$

Then, by monotonicity,  $\pi_0$  gives the share of always-takers ( $AT_D$ ), as  $E(D|X, Z_D = 0) = \pi_0$ . The share of compliers ( $C_D$ ) is  $\pi_1$ , reflecting individuals who change  $D$  when  $Z_D$  switches from 0 to 1. The share of never-takers ( $NT_D$ ), thus, is  $1 - \pi_0 - \pi_1$ .

Knowing these shares, we can estimate  $Y^1$  and  $Y^0$  for the compliers, to get the (local average) total treatment effect according to [Imbens and Rubin \(1997\)](#). This is possible with the following regression:

$$\begin{aligned} Y = & \delta_0 + \delta_1 [\mathbb{1}(D = 1)\mathbb{1}(Z_D = 0)] + \delta_2 [\mathbb{1}(D = 1)\mathbb{1}(Z_D = 1)] \\ & + \delta_3 [\mathbb{1}(D = 0)\mathbb{1}(Z_D = 0)] + \delta_4 [\mathbb{1}(D = 0)\mathbb{1}(Z_D = 1)] + \mathbf{X}'\boldsymbol{\delta}^X + v \end{aligned} \quad (10)$$

The coefficient  $\delta_1$  gives the always-taker-specific mean of  $Y$  (if the  $\mathbf{X}$  variables are not demeaned, the coefficients must be adjusted). Assumption 3 ensures that  $\delta_1 = E(Y^1|AT_D)$ . Likewise,  $\delta_2$  gives the mean for a subgroup formed of always-takers and (treated) compliers (that is, individuals with  $D = 1$  and  $Z_D = 1$ ). Vice versa,  $\delta_4$  is the mean of never-takers (thus,  $\delta_4 = E(Y^0|NT_D)$ ), whereas  $\delta_3$  is the mean of untreated compliers and never-takers. To estimate expected outcomes for treated and untreated compliers, we use  $\delta_2$  and  $\delta_3$  and adjust them for  $\delta_1$  and  $\delta_4$  with the help of the group shares. Thus,

$$\begin{aligned} E(Y^1|C_D) &= \frac{\delta_2(\pi_0 + \pi_1) - \delta_1\pi_0}{\pi_1} \\ E(Y^0|C_D) &= \frac{\delta_3(1 - \pi_0) - \delta_4(1 - \pi_0 - \pi_1)}{\pi_1}. \end{aligned}$$

---

<sup>19</sup>Of course, this can be estimated by two-stage least squares, and our approach delivers the algebraically same results.

The TTE is the difference between these quantities. Next, we do the same to get the  $E(Y^{jk}|C_D)$  and jointly address endogeneity in  $D$  and  $M$ . This means running Eq. (7) and Eq. (8) four times for all possible strata of  $D$ ,  $Z_D$  and determine  $MTR_{AT}^{11}$ ,  $MTR_{AT,CT}^{11}$ ,  $MTR_{AT}^{10}$ ,  $MTR_{AT,CT}^{10}$ ,  $MTR_{NT}^{01}$ ,  $MTR_{NT,CU}^{01}$ ,  $MTR_{NT}^{00}$ , and  $MTR_{NT,CU}^{00}$ . To reduce the notational burden,  $AT$  stands for always-takers of  $D$  and  $NT$  for never-takers of  $D$ .  $CT$  is treated compliers (compliers to  $Z_D$  and  $Z_D = 1$ ), and  $CU$  is untreated compliers (compliers to  $Z_D$  and  $Z_D = 0$ ).  $MTR_{AT,CT}^{11}$  is the MTR identified for the group of always-takers and compliers, and we follow the same approach as above to derive  $MTR_{CT}^{11}$ .

To ensure that Assumption 1 (randomization of  $Z_M$ ) holds even when conditioning on  $Z_D$ , we need to make an additional assumption:

*Assumption 4:* The two instruments  $Z_M$  and  $Z_D$  are conditionally independent.

This is necessary when stratifying the group of  $Z_D$  compliers further by their reaction to  $Z_M$ . We test one implication of this assumption later by evaluating whether the instrument  $Z_D$  gives us any predictive power to determine  $Z_M$ .

We could either purge the influence of the observable characteristics  $\mathbf{X}$  from the dependent variable by OLS or the estimator proposed by [Robinson \(1988\)](#). We opt for the former for convenience<sup>20</sup> and non-parametrically regress this cleaned variable on the propensity score at different evaluation points over the unit interval of the propensity score. With the local slope and the level coefficients at specific values of the propensity score over the unit interval, we can apply Eq. (7) and Eq. (8) to compute the eight MTRs formed by every combination of  $Z_D$ ,  $D$ , and  $M$ .

Again, holding fixed  $D$  and  $M$ , differences between the curve with  $Z_D = 1$  and the corresponding one with  $Z_D = 0$  must arise only because of the compliers if the typical LATE/IV assumptions hold. Because we know the share of compliers, we can apply the principle of [Imbens and Rubin \(1997\)](#) and calculate

$$\begin{aligned} MTR_C^{1k}(U_M = p) &= \frac{MTR_{AT,CT}^{1k}(U_M = p)(\pi_0 + \pi_1) - MTR_{AT}^{1k}(U_M = p)\pi_0}{\pi_1} \\ MTR_C^{0k}(U_M = p) &= \frac{MTR_{NT,CU}^{0k}(U_M = p)(1 - \pi_0) - MTR_{NT}^{0k}(U_M = p)(1 - \pi_0 - \pi_1)}{\pi_1} \\ &\quad \forall k \in \{0, 1\}. \end{aligned}$$

<sup>20</sup>To do this, we run, as a pre-processing step, the following regression where we restrict the control variables to have the same effect for all subgroups (the results are robust with respect to the chosen pre-processing method):

$$\begin{aligned} Y &= \alpha_{DZ_DM} + \beta_{DZ_DM}p + \mathbf{X}'\gamma + \mathbf{X}'p\delta + \varepsilon \\ \tilde{Y} &= Y - (\mathbf{X} - \bar{\mathbf{X}})'\hat{\gamma} - (\mathbf{X} - \bar{\mathbf{X}})'p\hat{\delta} \end{aligned}$$

Then, we compute an adjusted dependent variable  $\tilde{Y}$  in the equation above where the  $\mathbf{X}$  values are fixed at the respective mean values and using the estimated coefficients from this regression. The  $\alpha_{DZ_DM}$  and  $\beta_{DZ_DM}$  coefficients are specific levels and slopes for the propensity score, respectively, and may differ between the eight possible combinations of  $D$ ,  $M$ , and  $Z_D$ . This ensures that  $\tilde{Y}$  contains all marginal effects of  $M$  on  $Y$  in each  $D \times Z_D$  cell (which inform about the corresponding  $Z_D$  types) but without  $\mathbf{X}$  confounding these correlations. Hence, we can use  $\tilde{Y}$  as the dependent variable in eight non-parametric local linear regressions on the propensity score  $p$ .

Finally, we also make the following assumption, which is necessary to identify direct and indirect effects for the population of  $Z_D$ -compliers (and not just a subset, e.g., defined by  $Z_M$  compliers).

*Assumption 5:* Full support of  $PS_M$  by  $M = 1$  and  $M = 0$  within the group of compliers of  $D$ .

### Estimation of $\Pr(M^0)$ , $\Pr(M^1)$ and mediated outcomes $Y^{1M^1}$ , $Y^{1M^1}$ , $Y^{1M^1}$ , and $Y^{1M^1}$

The final missing pieces are  $E(M^1|U_M, C_D)$  and  $E(M^0|U_M, C_D)$  for the  $Z_D$  compliers and the mediated outcome curves (estimates of  $Y^{jM^l}$ ) that result from it. Regarding the former, we can simply compute  $E(\mathbb{1}[\Pr(M = 1|Z_M) \geq U_M^j]|U_M^j)$  along  $U_M$  in the four cells determined by  $D$  and  $Z_D$ , as outlined above, and, in the final step, apply the simple [Imbens and Rubin \(1997\)](#) formula to adjust the quantity to the compliers.

Now, we can identify the six quantities  $\text{MTR}_C^{jk}$  and  $E(M^j|U_M, C_D)$ , which are conditional expectations  $Y^{jk}$  and  $M^j$ , respectively, for the compliers along  $U_M$ . In the final step, we compute mediated outcomes  $Y^{jM^l}$  and determine the total, direct, and indirect effects. This is straightforward, as when applying  $E(\cdot|U_M, C_D)$  to the hypothetical observation rule  $Y^{jM^l} = Y^{j1}M^l + Y^{j0}(1 - M^l)$ , it shows that  $E(Y^{jM^l}|U_M, C_D)$  is a weighted mean of  $\text{MTR}_C^{j1}$  and  $\text{MTR}_C^{j0}$ , with weights determined by  $E(M^l|U_M, C_D)$ . Because the MTRs and  $E(M^j|U_M)$  are average outcomes along the margin of indifference, we can solve the identification problem inherent in  $Y^{1M^0}$  and  $Y^{0M^1}$ , which are never observed. This is because, at each value of  $U_M$ , compliers of  $D$  (our target group) are similar in their observable and unobservable characteristics (and  $D$  and  $M$  are randomized with perfect compliance). Hence, we know their outcome if they had chosen a different treatment and mediator state and can also disentangle the outcome effect from the mediator effect at this margin. This is why the MTE approach is so appealing for mediation analyses.

Standard errors of all effects are estimated by replicating the entire procedure 200 times on a random bootstrap sample clustered on country and birth year.

## 4.4 Simulation studies

In the Appendix, we try several different data-generating processes and show how our proposed solution identifies the actual effects in these DGPs while, for example, classical separate 2SLS fails when effect heterogeneity is assumed to be strong. We test situations with binary and continuous instruments, among others. We note that – for cases where neither treatments nor instruments are binary – the estimator by [Frölich and Huber \(2017\)](#) performs equally well (results not shown in the Appendix for the sake of brevity). Given that our estimator has the option to use stronger functional form assumptions, running time can be much shorter than using [Frölich and Huber \(2017\)](#), which may make applications with many control variables and large data sets more feasible (obviously at the cost of stronger assumptions).

## 5 Results

### 5.1 Evidence on the identifying assumptions and other threats to the validity of the estimates

Results from the IV validity tests in Table 4, columns (1)–(3) as well as absence of pre-trends Figure 3 (for effects of  $Z_D$  on  $Y$ ) and as shown in Schmitz and Westphal (2021) (for effects of  $Z_M$  on  $Y$ ) suggest that *Assumptions 1* and 3 (the LATE assumptions for labor force participation and education) hold. Decomposing education’s local average treatment effect on cognitive decline requires additional assumptions for the mediation analysis and other general threats. First, we assess the plausibility of additive separability (*Assumption 2*), which we need because our instrument for labor force participation is binary. This assumption implies an equal slope of the MTEs in every cell determined by the control variables. Although estimating MTEs in every subgroup is infeasible (therefore, we need to make this assumption), we can estimate different MTE curves in a few relevant subgroups of our data. As an arbitrary choice, we take gender and birth year and find the slopes of our MTEs almost equal (Figure A1). Second, we gauge the conditional independence of the instruments (*Assumption 4*). One testable implication of this assumption is that information on one instrument does not help predict the other. Table A4 suggests that this is the case: individuals affected by compulsory schooling cannot retire at different ages conditional on control variables compared to those not. Third, we can show that we have full support for the propensity score of  $M$  (*Assumption 5*), enabling the estimation of marginal treatment effects over the whole unit interval. Figure A2 depicts the resulting distribution of the propensity score by labor force status (complete estimation results are presented in Table A3). The common support ranges from 0.026 to 0.991—nearly the whole unit interval. Hence, the resulting TTE from our unconstrained MTE approach should recover the TTE estimated by 2SLS in Table 7.

Fourth, we want to scrutinize whether our setting suffers from problems in two-way fixed effects settings, as demonstrated, for instance, in Sun and Abraham (2021). The authors demonstrate that conventional event-study estimates could get contaminated with effects from other relative time periods. Figure A3 plots their proposed estimator’s results that prevent such contamination. It shows that accounting for this contamination, the first-stage and the reduced-form effects are somewhat larger. The difference is small, however, leaving the general conclusion unchanged. Moreover, the difference reduces for the IV estimates (the ratio of reduced form and first stage).

### 5.2 Estimated potential outcomes and treatment effects

Now we can turn to our main results—a formal mediation analysis of the effect of education on cognitive abilities that controls labor force participation as a potential mediator. To this end, we present the total treatment effect (TTE) and the two respective indirect and direct treatment effects in Table 7. Subsequently, we discuss the underlying components that



constitute these parameters (the MTR curves) in more detail. Turning to Table 7, the first line (labeled as MTE) is the main result of this paper. We contrast these estimates with those of three other estimators. First, we employ a slightly adjusted estimator of our MTE approach, where we estimate the ITEs as before but use 2SLS to estimate the TTE and then derive the DTEs using Eq. (1). By exploiting this restriction, we may gain efficiency in estimating the TTE and the DTEs. Second, we employ a conventional IV approach using IV versions of Eq. (2) and Eq. (3). Lastly, we ignore the endogeneity of  $D$  and  $M$  and estimate these parameters by OLS for completeness. Concerning the estimated results, the TTE amounts to 0.804–0.864 depending on the estimator (we ignore OLS because it is inconsistent). The difference in the TTE between the MTE and 2SLS estimator is remarkably small and negligible. This is reassuring that our estimator does well in estimating the local average treatment effect. A relevant difference is the larger standard errors due to a more inefficient estimation with many estimated parameters more than necessary for the LATE in three different regressions. Nonetheless, the MTE-TTE is still significant at the 10 percent level.

Table 7: Main results—total, direct, and indirect treatment effects

	Total treatment effect  $TTE = LATE$	Effect decomposition			
		Indirect TEs		Direct TEs	
		$ITE(1)$	$ITE(0)$	$DTE(1)$	$DTE(0)$
MTE	0.864* (0.505)	0.293* (0.153)	0.043 (0.052)	0.822* (0.494)	0.571 (0.488)
MTE (derived)	0.811** (0.380)	0.293* (0.153)	0.043 (0.052)	0.768** (0.377)	0.518 (0.395)
2SLS	0.804** (0.379)	0.184 (0.140)	−0.209 (0.164)	1.013** (0.485)	0.620* (0.354)
OLS	1.418*** (0.052)	0.039*** (0.005)	0.050*** (0.009)	1.368*** (0.054)	1.380*** (0.054)

Number of observations: 80,763. Control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects, and male. Bandwidth = 0.25. Bootstrap standard errors (200 replications) in parentheses clustered on the birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Turning to the effect decomposition, recall that by construction  $ITE(1)$  and  $DTE(0)$  as well as  $ITE(0)$  and  $DTE(1)$  each form a pair that each adds up to the total treatment effect.  $ITE(1)$  amounts to 0.293 for the MTE estimator (by definition, MTE and MTE (derived) are equivalent for the ITEs). Thus,  $ITE(1)$  amounts to more than one-third of the total effect and is significantly estimated at the 10 percent level. Estimating this effect by 2SLS yields a much smaller (probably biased) estimate. In our application,  $ITE(1)$  measures the contribution of a changed labor force participation caused by compulsory schooling for the more educated (treated) compliers. Education made these individuals work to older ages in their jobs (probably of a different quality than for less-educated compliers) before retiring.

This quantitatively meaningful parameter documents that the TTE is rather not constant over the life course but may arise through individual decisions after education is finished—in particular, decisions on the jobs and the retirement timing of these individuals.

The remaining part for the TTE is  $DTE(0)$ , i.e., the effect of increasing education due to compulsory schooling when labor force participation is fixed at  $M^0$ —the hypothetical labor force participation for compliers without the compulsory schooling reform. This effect amounts to 0.571 without a restriction and 0.518 when we use the TTE and  $ITE(1)$  to derive the DTE. Both effects are not statistically significant at any conventional level. However, the magnitude of the effect is not negligible. In any case, given that the DTE includes the direct effect of education plus all other mediating forces unrelated to labor force participation, it again outlines the relevance of labor force participation for the education effect.

Turning to the other pair,  $ITE(0)$  and  $DTE(1)$ , a slightly different picture emerges. We see that the indirect effect of labor force participation is much smaller (a contribution of 0.043 words to the TTE) and insignificant. Only slightly more than five percent of TTE is caused by the effect of retirement on the cognitive abilities of the compliers with less education. The difference between  $ITE(0)$  and  $ITE(1)$  is that the former is the effect of retirement on cognitive abilities for individuals with lower education (as opposed to the more-educated compliers for which the  $ITE(1)$  applies). This demonstrates an important complementary effect between education and labor force participation. Only more educated compliers gain through labor force participation, likely because their job environment is more stimulating. In contrast, less educated compliers do not gain through working to older ages. The lower  $ITE(0)$  mechanically forces  $DTE(1)$  to be larger than  $DTE(0)$ . To explain this result, consider the definition of  $DTE(1) = E(Y^{1M^1} - Y^{0M^1})$ . It is the causal effect of education on cognitive abilities if the individual retirement behavior is like that of the more educated individuals. Because more educated individuals retire later, they are longer in arguably more stimulating environments where they can sustain higher cognitive abilities more easily. Hence, this also shows the complementarity between education and labor force participation.

Now, we want to unravel our results by showing them along the margin of indifference for retirement (along the index  $U_M$ ) because this is key for our approach to work. To do this, we plot the four MTR functions—the average level of cognitive abilities by the education dummy  $D$  and the labor force dummy  $M$  along  $U_M$  for compulsory schooling compliers—in the left panel of Figure 4. For the sake of visibility, we ignore the standard errors of these lines. It shows that relative to the unconditional mean of the recall score in our sample ( $\bar{Y}_{it} = 10.41$ ), almost all MTRs for all values of  $U_M$  are lower. Only some compulsory schooling compliers with more education ( $D = 1$ ) and a low preference for retiring (high  $U_M$ ) who also do not retire ( $M = 0$ ) are still in the labor force and have higher cognitive scores than the overall average. This is unsurprising, as only compulsory schooling compliers at the lowest educational margins contribute to the plotted MTRs. Moreover, we can see that almost across the whole unit interval,  $E(Y^{10}|U_M = p, C_D)$  dominates the other MTRs. Among this group, only individuals with the highest retirement preference  $U_M < .2$  have one of the lowest average recall scores. The second highest ability curve is for individuals still in the labor force with years of schooling lower than the new compulsory standard



$(E(Y^{01}|U_M = p, C_D))$ . Given that both curves for labor force participating individuals are the highest, emphasize the role of work-related cognitive stimulation in maintaining cognitive abilities at older ages. The remaining two curves for individuals out of the labor force are lower. Somewhat surprisingly, perhaps, education does not seem to have a consistently positive effect for individuals who are retired (the difference between the red MTRs). This finding may suggest that employment and schooling are complementary for maintaining cognitive abilities, emphasizing the potential role of labor force participation as a mediating channel. We now focus on this indirect effect more formally.

The middle panel of Figure 4 informs about the retirement probabilities by education indicator  $D$  for compulsory schooling compliers ( $Pr(M^1 = 1|C_D, U_M)$  and  $Pr(M^0 = 1|C_D, U_M)$ ). These quantities both decline mechanically along  $U_M$ . It is also clearly visible that at all  $U_M$  values, the probability of being retired is lower for the more educated compliers. This means the indirect treatment effect comprises individuals with all possible retirement preferences.

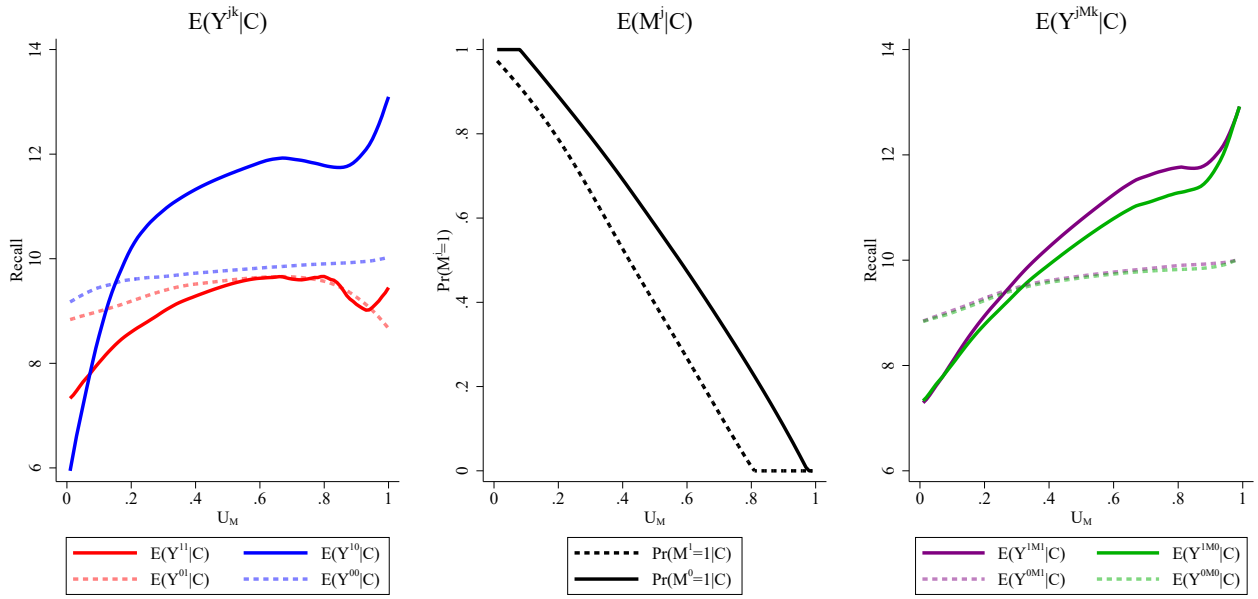
The last panel presents the mediated outcomes  $Y^{jM^j}$ , aggregated curves from the first two panels by weighting the four potential outcomes by the potential retirement probabilities for more and less educated compliers.<sup>21</sup> Average differences between any two of those lines give a certain mediation effect. For instance, the differences between the solid (dashed) purple and green lines yield the  $ITE(1)$  ( $ITE(0)$ ). The difference between the purple (green) solid and dashed line gives the  $DTE(1)$  ( $DTE(0)$ ). This figure, particularly the left panel, alludes to the primary cause of differences in the mediation effects. The high cognitive abilities of more educated compliers still in the labor force is causing the TTE in general, but also the differences between the ITEs, in particular. This finding applies to individuals with almost all retirement preferences—except for the very few with the highest preference. This emphasizes once again the complementarity between education and labor force participation.

Before we conclude, Table 8 documents the robustness of our estimates to some meaningful changes in the estimation procedure (i.e., the size of the bandwidth), the sample composition, and the retirement definition. Concerning the bandwidth in the nonparametric MTR estimation, the effects barely change quantitatively. It does not make a difference if we additionally instrument labor force participation by an indicator for the official retirement age (ORA), as is sometimes done in the literature on the effects of retirement. Dropping the unemployed, disabled, or homemaker increases all mediation parameters, but qualitatively, the results are the same (except now, the  $ITE(0)$  is also meaningful in magnitude). Changing the retirement definition to include the unemployed has no visible implications for the magnitude of the effects. Stratifying the effects by gender, however, outlines important heterogeneities in the effects. Whereas males have a smaller TTE than females, their  $ITE(1)$  is considerably larger. This may be because males typically have been more career-oriented, including working until statutory retirement age. When retiring, more will change for them, and the drop in the level of stimulation of their cognitive abilities may thus be larger. Education, in turn, may change more for females apart from their working environment, including

---

<sup>21</sup>The treatment-specific quantities are related as follows:  $Y^{jM^j} = Y^{j1}M^j + Y^{j0}(1 - M^j)$  for the observed and complier-specific outcome and  $Y^{jM^i} = Y^{j1}M^i + Y^{j0}(1 - M^i)$ , where  $j \neq 0$ , for the counterfactual complier-specific outcome if the labor force participation would be manipulated to the other treatment state.

Figure 4: Estimated marginal treatment response functions for compulsory-schooling compliers



Notes: Number of observations: 80,763. Control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects, and male.  $C$  represents compliers to the compulsory schooling reform (also denoted  $C_D$  in the text).

working at all (even if small hours), finding a different partner, etc. Finally, we scrutinize the results regarding differences in the compulsory schooling reforms. Keeping only countries where the compulsory schooling increase was one year may create a more homogeneous sample of countries. This attenuates the TTE somewhat and reduces the precision of all parameters, but the magnitude of all effects (except  $ITE(0)$ , as before) remains meaningful. In total, we confirm the robustness of our results in some important dimensions, suggesting that our results are not driven by some arbitrary choices in the sample selection, treatment definition, or estimation procedure. Although we lack the power to detect statistically significant effects at the 10 percent in all specifications and subsamples, the significance in the magnitude of the coefficients is unaffected, and many of the significant effects are very close to the 5 percent level, for instance, with a bandwidth of 0.3 or if unemployed are excluded.

## 6 Conclusion

We study the interaction of education in adolescence and labor force participation around retirement age and its effect on the cognitive abilities of individuals in Europe aged 55–70. Our main goal is to separate the total effect of education on older-age cognitive abilities into direct and indirect effects through labor force participation. By this, we aim at putting the results found in the literature so far in a more consistent perspective—as we believe that the age gradient in the effects (see introduction) may be caused by downstream differences in the cognitive environment (such as differences in labor force participation around retirement ages) long after education is finished. To this end, we conduct a causal mediation analysis. Since education and retirement are subject to individual choice (and, thus, endogenous),

Table 8: Robustness and other specifications

	Total treatment effect  <i>TTE = LATE</i>	Effect decomposition				<i>N</i>
		Indirect TEs		Direct TEs		
		<i>ITE</i> (1)	<i>ITE</i> (0)	<i>DTE</i> (1)	<i>DTE</i> (0)	
Bandwidth = 0.15	0.857* (0.466)	0.243 (0.179)	0.058 (0.055)	0.799* (0.455)	0.614 (0.459)	80,763
Bandwidth = 0.2	0.923* (0.480)	0.284* (0.161)	0.049 (0.054)	0.875* (0.468)	0.639 (0.467)	80,763
Bandwidth = 0.3	0.811 (0.501)	0.281* (0.147)	0.038 (0.050)	0.773 (0.489)	0.530 (0.478)	80,763
Bandwidth = 0.35	0.809* (0.485)	0.269* (0.144)	0.035 (0.050)	0.775 (0.474)	0.540 (0.462)	80,763
With ERA and ORA as instruments for <i>M</i>	0.936* (0.507)	0.241 (0.153)	0.018 (0.044)	0.918* (0.501)	0.695 (0.501)	80,763
Without unemployed, disabled, homemakers	1.132** (0.489)	0.398* (0.205)	0.153*** (0.048)	0.980** (0.475)	0.734* (0.417)	68,779
Unemployed = retired	0.967* (0.544)	0.290 (0.179)	0.009 (0.062)	0.958* (0.543)	0.677 (0.533)	80,763
Male	0.582 (0.659)	0.552** (0.261)	0.099 (0.0754)	0.483 (0.631)	0.0294 (0.612)	43,397
Female	1.148 (0.827)	0.365 (0.309)	0.0274 (0.0677)	1.121 (0.811)	0.783 (0.665)	37,366
One year increase in CS	0.729 (0.816)	0.365 (0.272)	0.003 (0.043)	0.726 (0.817)	0.364 (0.818)	59,635

Control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects and male. ERA (ORA) refer to indicators of being above the early (official) retirement age in the respective country. Bootstrap standard errors (200 replications) in parentheses clustered on birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

we exploit exogenous variation from compulsory schooling reforms and early retirement regulations for identification. We demonstrate how the marginal treatment effects framework can be used to conduct such a causal mediation analysis that accommodates heterogeneous treatment effects and non-compliance in educational and labor supply decisions.

We pool data from SHARE and ELSA on 80,000 observations from several European countries from 2002-2020. The data include experimentally collected measures of cognitive abilities (word recall test and verbal fluency test). In the first step, we can replicate the effect of education on cognitive abilities as found in the literature. When we split this effect into a direct effect of education and an indirect effect through labor force participation, we can show that retirement may be crucial for the onset of a cognitive decline. Retirement behavior

for more educated compulsory schooling compliers may explain more than one-third of the total effect. Moreover, we find evidence that schooling and labor supply are complementary, i.e., more education and labor force participation together seem to protect most against a cognitive decline. These results may explain the heterogeneous effect patterns that other studies found, which could be deemed inconsistent when not accounting for labor supply as a mediator.

Of course, later-life labor force participation is only one of an array of potential mediators of the effect of education on cognitive decline and not necessarily the most important one. Occupational choice and health behavior are probably two things that come to mind directly. Moreover, middle-life education, middle-life labor force participation, family status, and other forms of cognitive stimulation are likely to play a role. Thus, even if claimed to be a causal mediation analysis, this analysis can only inform about a small detail of the bigger picture of “cognitive decline”. What is called “direct effect” is, as obviously in all mediation analyses, only a compound one of the actual direct effects and other unmeasured indirect effects. Future work might simultaneously consider more indirect paths of education on cognitive abilities—with the increased demand for data and exogenous variation that comes along with this.

Nonetheless, important policy implications may arise. The policy could act today to reap the effects of past education reforms by enabling individuals to maintain a more cognitively stimulating environment until older ages. More liberal retirement policies and flexible work arrangements appear to be important instruments for this, and they would come at almost no cost. Our results also demonstrate important side effects of education, which are not detectable shortly after education is completed but emerge over the life course. This life-course perspective needs to be considered when assessing the non-monetary benefits of education.

## Acknowledgments

SHARE: This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6, 7 and 8 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, 10.6103/SHARE.w7.711, 10.6103/SHARE.w8.100, 10.6103/SHARE.w8ca.100), see [Börsch-Supan et al. \(2013\)](#) for methodological details. The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01\_AG09740-13S2, P01\_AG005842,

P01\_AG08291, P30\_AG12815, R21\_AG025169, Y1-AG-4553-01, IAG\_BSR06-11, OGHA\_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see [www.share-project.org](http://www.share-project.org)).

ELSA: This analysis uses data or information from the Harmonized ELSA dataset and Codebook, Version E as of April 2017 developed by the Gateway to Global Aging Data. The development of the Harmonized ELSA was funded by the National Institute on Aging (R01 AG030153, RC2 AG036619, 1R03AG043052). For more information, please refer to [www.g2aging.org](http://www.g2aging.org).

## References

- American Psychological Association (1995). Intelligence: Knowns and Unknowns, Report of a task force convened by the American Psychological Association. *Science Directorate, Washington DC [LGH]*.
- Arendt, J. N. (2005). Does education cause better health? a panel data analysis using school reforms for identification. *Economics of Education Review*, 24(2):149–160.
- Atalay, K., Barrett, G. F., and Staneva, A. (2019). The effect of retirement on elderly cognitive functioning. *Journal of Health Economics*, 66:37 – 53.
- Banks, J. and Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *The Economic Journal*, 122(560):418–448.
- Banks, J., O’Dea, C., and Oldfield, Z. (2010). Cognitive Function, Numeracy and Retirement Saving Trajectories. *The Economic Journal*, 120(548):381–410.
- Banks, J. and Oldfield, Z. (2007). Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving. *Fiscal Studies*, 28(2):143–170.
- Bonsang, E., Adam, S., and Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of Health Economics*, 31(3):490–501.
- Börsch-Supan, A. (2019a). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.700. Technical report.
- Börsch-Supan, A. (2019b). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.700. Technical report.
- Börsch-Supan, A. (2019c). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 - SHARELIFE. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.700. Technical report.
- Börsch-Supan, A. (2019d). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.700. Technical report.
- Börsch-Supan, A. (2019e). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.700. Technical report.
- Börsch-Supan, A. (2019f). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.700. Technical report.
- Börsch-Supan, A. (2019g). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w7.700. Technical report.
- Börsch-Supan, A. (2021). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 1.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8.100. Technical report.
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, Sabrina, o. b. o. t. S. C. C. T. (2013). Data Resource Profile: The Survey

- of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42(4):992–1001.
- Börsch-Supan, A. and Jürges, H. (2005). The Survey of Health, Aging and Retirement in Europe – Methodology. Technical report, Mannheim: Mannheim Research Institute for the Economics of Aging.
- Brinch, C. N. and Galloway, T. A. (2012). Schooling in adolescence raises iq scores. *Proceedings of the National Academy of Sciences*, 109(2):425–430.
- Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.
- Brugiavini, A., Orso, C. E., Genie, M. G., Naci, R., and Pasini, G. (2019). Combining the retrospective interviews of wave 3 and wave 7: the third release of the share job episodes panel. share working paper series: 36-2019. munich: Mea, max planck institute for social law and social policy. Technical report.
- Brunello, G., Fabbri, D., and Fort, M. (2013). The Causal Effect of Education on Body Mass: Evidence from Europe. *Journal of Labor Economics*, 31(1):195–223.
- Brunello, G., Fort, M., and Weber, G. (2009). Changes in compulsory schooling, education and the distribution of wages in europe. *The Economic Journal*, 119(536):516–539.
- Brunello, G., Weber, G., and Weiss, C. T. (2016). Books are Forever: Early Life Conditions, Education and Lifetime Earnings in Europe. *The Economic Journal*, 127(600):271–296.
- Carlsson, M., Dahl, G. B., Öckert, B., and Rooth, D.-O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, 97(3):533–547.
- Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81.
- Carneiro, P. and Lee, S. (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*, 149(2):191–208.
- Celidoni, M., Bianco, C. D., and Weber, G. (2017). Retirement and cognitive decline. a longitudinal analysis using share data. *Journal of Health Economics*, 56:113 – 125.
- Chandra, A., Coile, C., and Mommaerts, C. (2023). What Can Economics Say About Alzheimer’s Disease? *Journal of Economic Literature*, 61(2):428–470.
- Chen, S. H., Chen, Y.-C., and Liu, J.-T. (2019). The Impact of Family Composition on Educational Achievement. *Journal of Human Resources*, 54(1):122–170.
- Chen, Y.-T., Hsu, Y.-C., and Wang, H.-J. (2020). A stochastic frontier model with endogenous treatment status and mediator. *Journal of Business & Economic Statistics*, 38(2):243–256.
- Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.
- Coe, N. B., von Gaudecker, H., Lindeboom, M., and Maurer, J. (2012). The Effect Of Retirement On Cognitive Functioning. *Health Economics*, 21(8):913–927.
- Cornelissen, T. and Dustmann, C. (2019). Early School Exposure, Test Scores, and Noncognitive Outcomes. *American Economic Journal: Economic Policy*, 11(2):35–63.
- Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409.
- De Chaisemartin, C. and d’Haultfoeuille, X. (2020). Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–96.
- De Chaisemartin, C. and D’Haultfoeuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.
- Dippel, C., Ferrara, A., and Heblich, S. (2020). Causal mediation analysis in instrumental-variables regressions. *Stata Journal*, 20(3):613–626.
- Fratiglioni, L. and Wang, H.-X. (2007). Brain Reserve Hypothesis in Dementia. *Journal of Alzheimer’s Disease*, 12(1):11–22.
- Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects—causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series*

- B (Statistical Methodology)*, 79(5):1645–1666.
- Garrouste, C. (2010). *100 Years of Educational Reforms in Europe: a contextual database*. PhD thesis, European Commission’s Joint Research Centre (JRC).
- Glymour, M. M., Kawachi, I., Jencks, C. S., and Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, 62(6):532–537.
- Gorman, E. et al. (2017). Schooling, Occupation and Cognitive Function: Evidence from Compulsory Schooling Laws. *SocArXiv*. October, 23.
- Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, 13(7):491–505.
- Hampf, F. (2019). The effect of compulsory schooling on skills: Evidence from a reform in Germany. Technical report, Ifo Working Paper.
- Heckman, J. J. (2007). The Economics, Technology, and Neuroscience of Human Capability Formation. *Proceedings of the National Academy of Sciences*, 104(33):13250–13255.
- Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, 46(3):289–324.
- Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic literature*, 48(2):356–398.
- Heckman, J. J. and Vytlacil, E. J. (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734.
- Hollenbach, J., Schmitz, H., and Westphal, M. (2024). Gene-environment interactions with essential heterogeneity. *Ruhr Economic Papers* (1105), RWI.
- Hong, G. (2010). Ratio of mediator probability weighting for estimating natural direct and indirect effects. *Proceedings of the Biometrics Section – American Statistical Association*, pages 2401–2415.
- Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics*, 29(6):920–943.
- Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71.
- Imbens, G. W. and Rubin, D. B. (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *The Review of Economic Studies*, 64(4):555–574.
- Kamhöfer, D. A. and Schmitz, H. (2016). Reanalyzing zero returns to education in germany. *Journal of Applied Econometrics*, 31(5):912–919.
- Kaufman, A. S. and Horn, J. L. (1996). Age changes on tests of fluid and crystallized ability for women and men on the kaufman adolescent and adult intelligence test (kait) at ages 17–94 years. *Archives of Clinical Neuropsychology*, 11(2):97–121.
- Keele, L., Tingley, D., and Yamamoto, T. (2015). Identifying mechanisms behind policy interventions via causal mediation analysis. *Journal of Policy Analysis and Management*, 34(4):937–963.
- Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, 346(6209):572–578.
- Lundborg, P., Nilsson, A., and Rooth, D.-O. (2014). Parental education and offspring outcomes: Evidence from the swedish compulsory school reform. *American Economic Journal: Applied Economics*, 6(1):253–78.
- Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710.
- Mazzonna, F. and Peracchi, F. (2017). Unhealthy Retirement? *Journal of Human Resources*, 52(1):128–151.
- Mogstad, M., Santos, A., and Torgovitsky, A. (2018). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, 86(5):1589–1619.
- Mourifié, I. and Wan, Y. (2017). Testing local average treatment effect assumptions. *The Review of Economics and Statistics*, 99(2):305–313.
- Nicoletti, C., Salvanes, K. G., and Tominey, E. (2023). Mothers Working during Preschool Years and Child Skills: Does Income Compensate? *Journal of Labor Economics*, 41(2):389–429.



- Nybom, M. (2017). The distribution of lifetime earnings returns to college. *Journal of Labor Economics*, 35(4):903–952.
- Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, page 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Ritchie, S. J. and Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-analysis. *Psychological Science*, 29(8):1358–1369.
- Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931–954.
- Rohwedder, S. and Willis, R. J. (2010). Mental Retirement. *Journal of Economic Perspectives*, 24(1):119–38.
- Rose, E. K. and Shem-Tov, Y. (2021). How Does Incarceration Affect Reoffending? Estimating the Dose-Response Function. *Journal of Political Economy*, 129(12):3302–3356.
- Salm, M., Siflinger, B., and Xie, M. (2021). The effect of retirement on mental health: Indirect treatment effects and causal mediation. Workingpaper, CentER, Center for Economic Research. CentER Discussion Paper Nr. 2021-012.
- Schiele, V. and Schmitz, H. (2023). Understanding cognitive decline in older ages: The role of health shocks. *European Economic Review*, 151:104320.
- Schmitz, H. and Westphal, M. (2021). The dynamic and heterogenous effect of retirement on cognitive decline. *Ruhr Economic Papers* 918, RWI.
- Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, 51(2):619–643.
- Seblova, D., Fischer, M., Fors, S., Johnell, K., Karlsson, M., Nilsson, T., Svensson, A. C., Lövdén, M., and Lager, A. (2021). Does Prolonged Education Causally Affect Dementia Risk When Adult Socioeconomic Status Is Not Altered? A Swedish Natural Experiment in 1.3 Million Individuals. *American Journal of Epidemiology*, 190(5):817–826.
- Smith, J. P., McArdle, J. J., and Willis, R. (2010). Financial Decision Making and Cognition in a Family Context. *The Economic Journal*, 120(548):363–380.
- Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., and Weir, D. R. (2014). Cohort Profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, 43(2):576–585.
- Steptoe, A., Breeze, E., Banks, J., and Nazroo (2003). Cohort profile: the English longitudinal study of ageing. *International Journal of Epidemiology*, 42(6):1640–1648.
- Strittmatter, A., Sunde, U., and Zegners, D. (2020). Life cycle patterns of cognitive performance over the long run. *Proceedings of the National Academy of Sciences*, 117(44):27255–27261.
- Sun, L. (2021). eventstudyinteract: interaction weighted estimator for event study.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Tchetgen, E. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *Annals of Statistics*, 40(3):1816–1845.
- Tymula, A., Rosenberg Belmaker, L. A., Ruderman, L., Glimcher, P. W., and Levy, I. (2013). Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, 110(42):17143–17148.
- van Kippersluis, H., O'Donnell, O., and van Doorslaer, E. (2011). Long-run returns to education: Does schooling lead to an extended old age? *Journal of Human Resources*, 46(4):695–721.
- Westphal, M., Kamhöfer, D. A., and Schmitz, H. (2022). Marginal college wage premiums under selection into employment. *The Economic Journal*, 132(646):2231–2272.



# Appendix

## A Additional tables and figures

Table A1: Number of observations

Country	Total	CS = 0		CS = 1		D = 0		D = 1	
		ERA = 0	ERA = 1	ERA = 0	ERA = 1	M = 0	M = 1	M = 0	M = 1
Austria	5705	473	2031	1548	1653	197	931	1868	2709
Germany	8273	2052	3332	1929	960	223	381	4179	3490
Spain	6764	2673	4091	0	0	1099	1401	2335	1929
Italy	2146	54	536	395	1161	78	207	1154	707
France	9799	1838	4445	2214	1302	547	1351	3958	3943
Greece	2419	1020	1399	0	0	431	281	1168	539
Czech Republic	3480	0	251	652	2577	2	12	1979	1487
England	42177	25827	16350	0	0	6409	10169	14081	11518
Total	80763	33937	32435	6738	7653	8986	14733	30722	26322

Notes: Working sample from SHARE and ELSA data.

Table A2: Regression results: total treatment effect for years of education

	Treatment: More education			Treatment: Years of education		Reduced Form
	OLS (1)	First stage (2)	2SLS (3)	First stage (4)	2SLS (5)	
More education ( $D$ )	1.423*** (0.052)		0.811*** (0.306)			
Years of education					0.997** (0.457)	
Post CS-reform ( $Z_1$ )		0.244*** (0.020)		0.199** (0.084)		0.198** (0.081)
Control variables	yes	yes	yes	yes	yes	yes

Notes: Number of observations in each regression: 80,763. Additional control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects and male. Standard errors in parentheses clustered on birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A3: Propensity Score ( $P(Z_M)$ ) estimation

	(1)		(2)	
	Coefficient	Standard Error	Marginal Effect	Standard Error
Above ERA ( $Z_M$ )	0.285***	(0.0161)	0.0869***	(0.0048 ¿ 7)

<hr/>				
$D \times Z_D$				
1 × 1	-0.345***	(0.0231)	-0.105***	(0.00703)
0 × 1 (empty, no never takers)				
1 × 0	-0.258***	(0.0117)	-0.0788***	(0.00355)
Country				
Germany	-0.104	(0.381)	-0.219***	(0.00748)
Spain	-0.193	(0.416)	-0.176***	(0.00830)
Italy	-1.911**	(0.718)	-0.137***	(0.00992)
France	-1.603***	(0.381)	-0.0806***	(0.00672)
Greece	1.919***	(0.465)	-0.230***	(0.0117)
Czech Republic	-5.177***	(0.561)	-0.134***	(0.00795)
England	1.018**	(3.09)	0.311**	(3.09)
Country-specific age trends				
Germany	-0.0108	(0.006)	-0.00329	(0.002)
Italy	0.0237*	(1.98)	0.00724*	(1.98)
France	0.0220***	(0.006)	0.00673***	(0.002)
Greece	-0.0446***	(0.007.76)	-0.0136***	(0.024)
Czech Republic	0.0782***	(0.009)	0.0239***	(0.285)
England	-0.0289***	(0.005)	-0.00881***	(0.002)
Age	0.0393***	(0.011)	0.0120***	(0.003)
Birth year				
1931	0.396	(0.323)	0.0362	(0.0354)
1932	-0.0255	(0.298)	-0.00307	(0.0355)
1933	0.376	(0.301)	0.0349	(0.0348)
1934	0.412	(0.297)	0.0373	(0.0350)
1935r	0.0452	(0.285)	0.00521	(0.0338)
1936	-0.0894	(0.282)	-0.0112	(0.0334)
1937	-0.112	(0.283)	-0.0142	(0.0333)
1938	-0.331	(0.284)	-0.0480	(0.0331)
1939	-0.438	(0.286)	-0.0676*	(0.0332)
1940	-0.668*	(0.289)	-0.116***	(0.0331)
1941	-0.762**	(0.292)	-0.139***	(0.0332)
1942	-0.948**	(0.295)	-0.188***	(0.0332)
1943	-1.045***	(0.299)	-0.216***	(0.0334)
1944	-1.104***	(0.304)	-0.233***	(0.0338)
1945	-1.318***	(0.308)	-0.301***	(0.0345)
1946	-1.413***	(0.313)	-0.331***	(0.0353)
1947	-1.487***	(0.318)	-0.356***	(0.0366)
1948	-1.618***	(0.324)	-0.400***	(0.0380)
1949	-1.741***	(0.329)	-0.442***	(0.0398)
1950	-1.862***	(0.335)	-0.483***	(0.0416)
1951	-1.917***	(0.341)	-0.502***	(0.0437)
1952	-2.125***	(0.347)	-0.570***	(0.0454)
1953	-2.191***	(0.354)	-0.591***	(0.0475)
1954	-2.291***	(0.361)	-0.621***	(0.0492)
1955	-2.469***	(0.367)	-0.672***	(0.0501)
1956	-2.566***	(0.374)	-0.698***	(0.0512)
Wave				
2	0.308***	(0.0332)	0.0869***	(0.00796)
3	0.402***	(0.0513)	0.115***	(0.0128)
4	0.576***	(0.0684)	0.168***	(0.0173)
5	0.749***	(0.0883)	0.221***	(0.0231)
6	0.891***	(0.108)	0.264***	(0.0288)
<hr/>				

7	1.292***	(0.129)	0.380***	(0.0331)
8	1.218***	(0.153)	0.360***	(0.0402)
Repetitions				
2	-0.190***	(0.0175)	-0.0579***	(0.00527)
3	-0.152***	(0.0196)	-0.0462***	(0.00592)
4	-0.181***	(0.0215)	-0.0551***	(0.00647)
5	-0.165***	(0.0263)	-0.0500***	(0.00794)
6	-0.0401	(0.0326)	-0.0122	(0.00987)
7	-0.137**	(0.0432)	-0.0416**	(0.0131)
8	0.147*	(0.0584)	0.0443*	(0.0175)
Male	-0.335***	(0.0109)	-0.102***	(0.00327)
Intercept	-0.336	(0.834)		
Observations	76379	76379		

Standard errors in parentheses

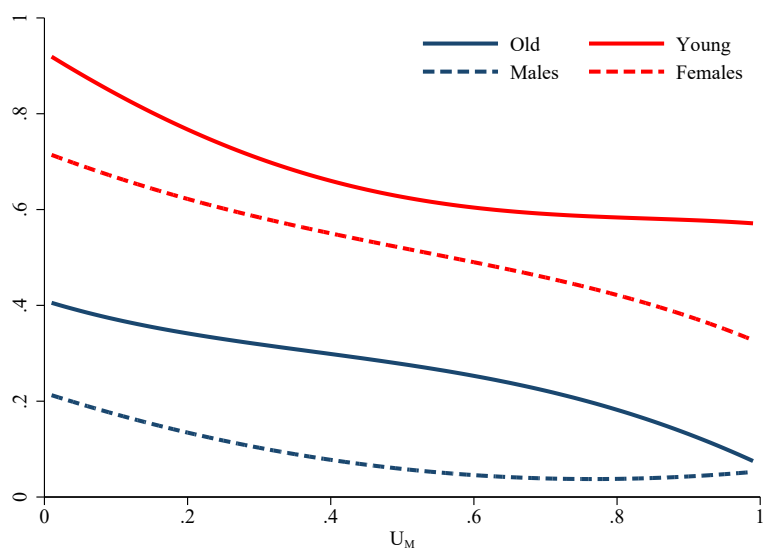
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table A4: Regression results: predicting  $Z_M$  with  $Z_D$  (conditional on controls)

	Dependent variable $Z_M$	
	Coefficient	Standard error
Compulsory schooling indicator $Z_D$ :	0.026	(0.025)
Control variables	yes	

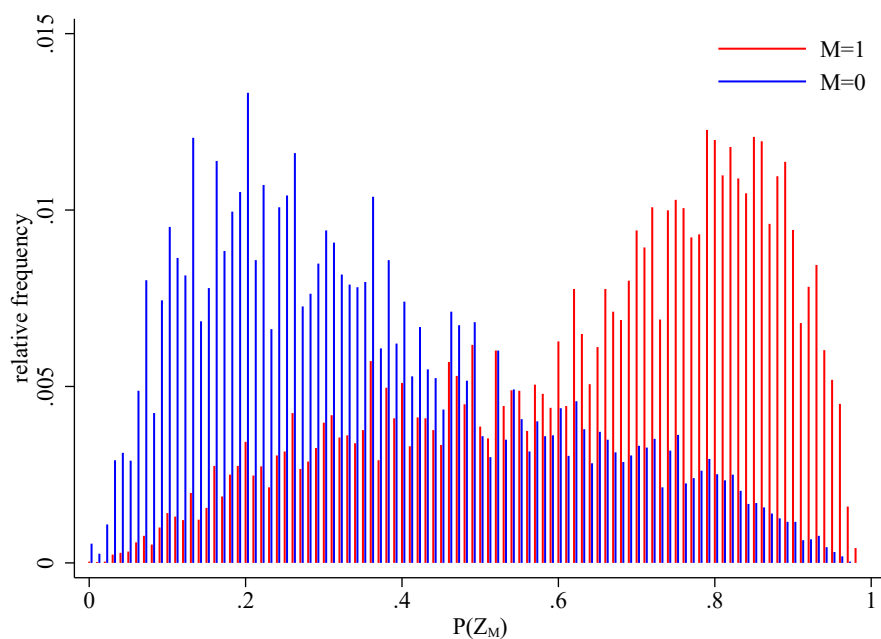
Notes: Number of observations: 80,763. Regression equation reads  $Z_M = \alpha + \beta Z_D + \mathbf{X}'\boldsymbol{\delta} + \varepsilon$ . Control variables are birth year fixed effects, interview wave fixed effects, country fixed effects, country-specific linear age trends, test repetition fixed effects and male. Standard errors in parentheses clustered on birth year-country level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.05$ .

Figure A1: Similarity of MTE slopes in relevant subgroups



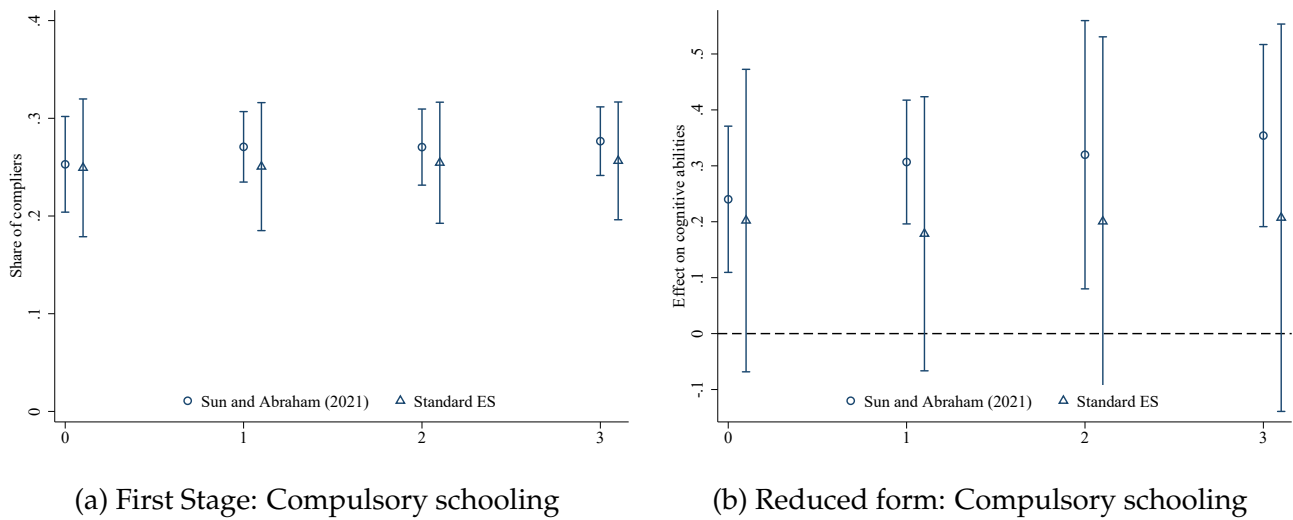
Notes: This graph plots MTE curves of labor force participation ( $M$ ) on recall for four subgroups (males, female, and old and young as determined by a median split).

Figure A2: Support of  $P(Z_M)$  by labor-force status ( $M$ )



Notes: This graph plots the relative frequency of the  $P(Z_M)$  values to lie in 0.01 bins of the propensity score by labor force status. The complete regression results are presented in Table A3.

Figure A3: Interaction-weighted event-study estimates (Sun and Abraham, 2021)



Notes: This figure contrasts the conventional event-study results from Figure 3 with results from the interaction-weighted estimator suggested by (Sun and Abraham, 2021). Event time is birth cohort minus pivotal cohort. Here, we use the Stata command `eventstudyinteract`, see Sun (2021).

## B Simulation studies

We carry out three different simulation studies to compare the performance of our estimator with a classic instrumental variables approach. Data generating process 1 (**DGP 1**) is the following simple case without control variables and a continuous instrument  $Z_2$ :

$$\begin{aligned}
Y^{11} &= \alpha_D + \alpha_M + \alpha_{DM} + U^{11} \\
Y^{10} &= \alpha_D + U^{10} \\
Y^{01} &= \alpha_M + U^{01} \\
Y^{00} &= U^{00} \\
M^1 &= \mathbb{1}(\gamma_Z^1 Z_2 + \alpha + U_M^1 > 0) \\
M^0 &= \mathbb{1}(\gamma_Z^0 Z_2 + U_M^0 > 0) \\
D &= \mathbb{1}(\beta_Z^D Z_1 + V > 0) \\
Z_1 &= \mathbb{1}(P > 0) \\
Z_2 &= Q \\
M &= M^1 D + M^0 (1 - D) \\
Y &= D M Y^{11} + D (1 - M) Y^{10} + (1 - D) M Y^{01} + (1 - D) (1 - M) Y^{00}
\end{aligned}$$

$P, Q$  independent standard normal. The parameters are set to  $\alpha_D = 1, \alpha_M = 1, \alpha_{DM} = 5, \alpha = 1, \beta_Z^D = 1, \gamma_Z^1 = 1, \gamma_Z^0 = 1$ . Moreover, the error correlation structure is as follows:

$$\begin{bmatrix} U^{11} \\ U^{10} \\ U^{01} \\ U^{00} \\ U_M^1 \\ U_M^0 \\ V \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 0.32 \\ 0.32 & 1 & 0.32 & 0.32 & 0.32 & 0.32 & 0.32 \\ 0.32 & 0.32 & 1 & 0.32 & 0.32 & 0.32 & 0.32 \\ 0.32 & 0.32 & 0.32 & 1 & 0.32 & 0.32 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 1 & 0.32 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 1 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 1 \end{bmatrix} \right)$$

All error terms are correlated; hence,  $D$  and  $M$  are endogenous. DGP 1 generates a complier share of 34% and a correlation of  $U_D$  and  $U_M$  of 0.32. We vary the error term structure in two ways:

**DGP 1'** (stronger correlation):

$$\begin{bmatrix} U^{11} \\ U^{10} \\ U^{01} \\ U^{00} \\ U_M^1 \\ U_M^0 \\ V \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 1 \end{bmatrix} \right)$$

**DGP 1''** (No correlation between  $U^{11}, U^{10}, U^{01}$ , and  $U^{00}$ , and no correlation between  $U_M^1$  and  $U_M^0$ ):

$$\begin{bmatrix} U^{11} \\ U^{10} \\ U^{01} \\ U^{00} \\ U_M^1 \\ U_M^0 \\ V \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 & 0.32 & 0.32 & 0.32 \\ 0 & 1 & 0 & 0 & 0.32 & 0.32 & 0.32 \\ 0 & 0 & 1 & 0 & 0.32 & 0.32 & 0.32 \\ 0 & 0 & 0 & 1 & 0.32 & 0.32 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 1 & 0 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 0 & 1 & 0.32 \\ 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 0.32 & 1 \end{bmatrix} \right)$$

### Estimation and results:

Figure B1, Figures S1 and S2, and Table S1 in the Supplementary Materials report the results using 200 simulation rounds with 50,000 observations per round. The two horizontal lines in each subfigure denote the (unconditional) indirect and direct effects, both for the full sample and for the subsample of compliers to instrument  $Z_1$ . We aim to estimate the effects for the compliers and compare the average estimated effects to the average true effects. The bars are averages over the 200 estimates gained by four different estimation procedures. While Figure B1 shows results for DGP 1, Figures S1 and S2 show results for DGP 1' and DGP 1'', respectively.

As a benchmark, we carry out OLS regressions where the parameters of the following two equations are estimated:

$$\begin{aligned} Y &= \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon \\ M &= \gamma_0 + \gamma_D D + \gamma_X X + u \end{aligned}$$

The mediation effects are estimated to be

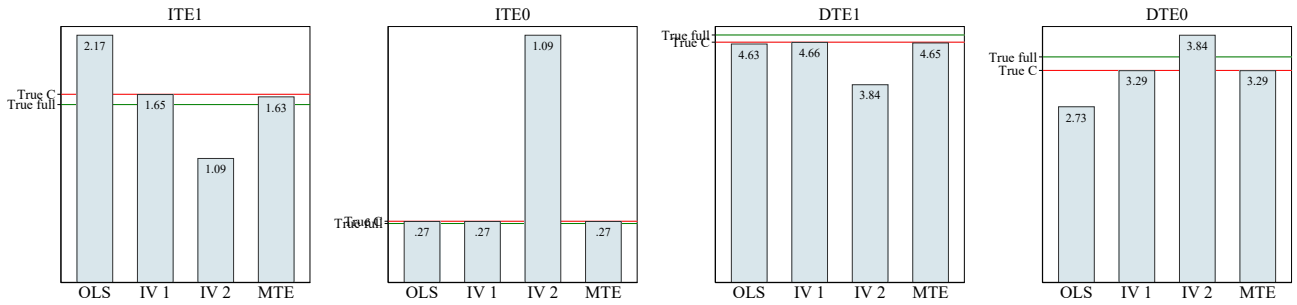
$$\begin{aligned} ITE(1) &= \gamma_D \cdot (\delta_M + \delta_{DM}) \\ ITE(0) &= \gamma_D \cdot \delta_M \\ DTE(1) &= \delta_D + \delta_{DM} \cdot (\gamma_0 + \gamma_D) \\ DTE(0) &= \delta_D + \delta_{DM} \cdot \gamma_0, \end{aligned}$$

and TTE (not reported in the figures but in the table) is derived using Eq. (1).

The results for the bars denoted by IV 1 in Figure B1 are generated by the same procedure with the only difference that the two equations above are estimated by two-stage least squares. Here,  $Z_1$  is used as an instrument for  $D$ ,  $Z_2$  is the instrument for  $M$ , and the interaction of  $Z_1$  and  $Z_2$  is used as an additional instrument in the equation with the interaction of  $D$  and  $M$ . Following Frölich and Huber (2017), we also present results when possible interaction effects of  $D$  and  $M$  are ignored in the IV estimations as sometimes seen in the literature. This means to assume  $ITE(1) = ITE(0)$  and  $DTE(1) = DTE(0)$ . The results are reported by the bars denoted IV 2. Finally, we report the results of our proposed estimator in the bars denoted by MTE.

The average effects (and, thus, biases when compared to the true effects) are reported in the figures, while Table S1 in the Supplementary Materials also reports the root mean squared error in the columns denoted by RMSE. DGP 1 shows that *OLS* produces biased estimates, particularly in estimating *ITE(1)* and *DTE(0)*. For instance, while the average true *ITE(1)* for compliers in DGP 1 is 1.65, *OLS* estimates it to be 2.17. The bias is aggravated when the errors are correlated more strongly, as in DGP 1' (see Figure S1). Likewise, a wrongly specified instrumental variables estimation (as in *IV 2*) yields estimates of *ITE* and *DTE* far away from the true effects. However, the more flexible *IV* specification with an interaction of *D* and *M* in *IV 1* only generates a negligible bias, where all average estimates are very close to the true average effects for the group of compliers. The same holds for our proposed estimator, as seen in line *MTE*. *IV 1* slightly dominates *MTE*, also in terms of RMSE, but, overall, the differences are negligible.

Figure B1: Results of DGP 1



Notes: Own calculations. This Figure plots the average results of DGP 1 after replicating the estimation respective procedures 200 times. Each panel refers to a specific indirect of direct mediation effect. The green and red horizontal line depict the average true effect for the whole sample and for the  $Z_1$  compliers only (our target parameter), respectively. *OLS* and *IV1* refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by *OLS* and two-stage least squares (2SLS), respectively. *IV2* reports the implied effect when no interaction term is used in the 2SLS outcome regression. *MTE* refers to the results of the estimation procedure derived in the paper. Full results including RMSE are reported in Table S1.

The results for DGP 1' and DGP 1'' are very similar and confirm that our method works also with different structures of error correlation.

## DGP 2: Binary instruments and control variables

DGP 2 shows that our approach works with binary instruments ( $Z_1$  and  $Z_2$ ) and control variables. The DGP is in the spirit of the one proposed by Frölich and Huber (2017) but adapted to a binary instrument  $Z_2$ .

$$\begin{aligned}
 Y &= 5D + 10M + 7DM + 5X + U \\
 M &= 1(-1.5 + 4Z_2 + 1D + 0.5X + V > 0) \\
 D &= 1(-1.5 + 2Z_1 + 0.5X + W > 0) \\
 Z_1 &= 1(0.5X + P > 0) \\
 Z_2 &= 1(Q > 0)
 \end{aligned}$$

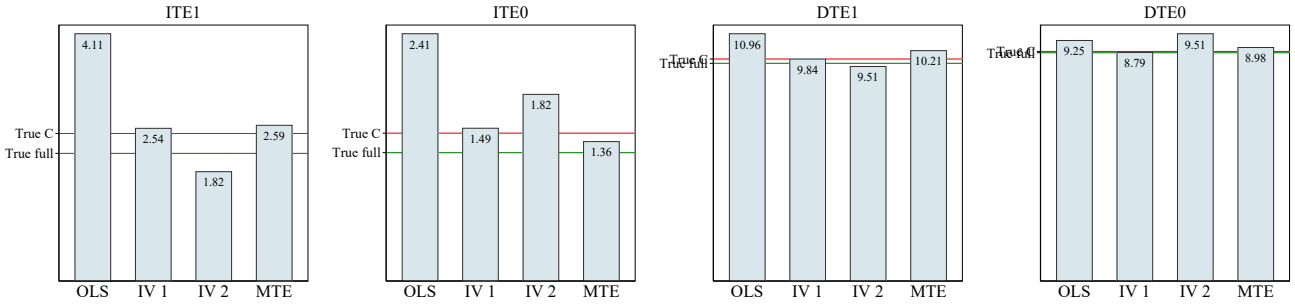


and

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} \sim \mathcal{N} \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix} \right]$$

$X, P, Q$  independent standard normal. Results are shown in Figure B2. Again, *MTE* and *IV 1* are the two methods that yield estimates close to the true ones. Here, *MTE* has the larger RMSE. Nevertheless, across all different types of DGPs, *MTE* is the most flexible procedure that allows the identification of effects even in situations when classic IV fails.

Figure B2: Results of DGP 2



Notes: Own calculations. This Figure plots the average results of DGP 2 after replicating the respective estimation procedures 200 times. Each panel refers to a specific indirect or direct mediation effect. The green and red horizontal lines depict the average true effect for the whole sample and the  $Z_1$  compliers only (our target parameter), respectively. OLS and IV1 refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by OLS and two-stage least squares (2SLS), respectively. IV2 reports the implied effect when no interaction term is used in the 2SLS outcome regression. MTE refers to the results of the estimation procedure derived in the paper. Full results, including RMSE, are reported in Table S1.

### DGP 3: Strong heterogeneity between complier groups

While DGP 1 and DGP 2 generate a fairly homogeneous data set—the effects do not vary strongly between complier types—we organize DGP 3 in a way that it produces very heterogeneous effects among complier types as well as regarding ITE(1) vs. ITE(0) and DTE(1) vs. DTE(0) in the group of compliers. Additionally, DGP includes a control variable that affects all other observables plus the unobserved complier types. To this end, we define three different DGPs for the three groups (compliers, always takers, and never takers) and then pool the data. In this sense, DGP 3 is more general than DGP 1 and 2 because we not only allow the three groups to have different values of unobservables. In addition, we let the unobservables have an arbitrarily different impact on the outcome across the three groups (holding selection on unobservables constant). Thus, this DGP is, in principle more realistic than DGP 1 and 2. The three subsamples are generated as follows:

Complier	Always taker	Never taker
$Y^{11} = -5 + 5X + 2U^{11}$	$Y^{11} = -10 + 5X - 2U^{11}$	
$Y^{10} = 5 + 5X + 4U^{10}$	$Y^{10} = -15 + 5X - 4U^{10}$	
$Y^{01} = -15 + 5X - 2U^{01}$		$Y^{01} = -15 + 5X - 2U^{01}$
$Y^{00} = -5 + 5X - 2U^{00}$		$Y^{00} = -5 + 5X - 4U^{00}$
$M^1 = \mathbb{1}(3.2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^1)$	$M^1 = \mathbb{1}(3.2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^1)$	$M^1 = \mathbb{1}(3.2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^1)$
$M^0 = \mathbb{1}(2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^0)$	$M^0 = \mathbb{1}(2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^0)$	$M^0 = \mathbb{1}(2 - \gamma_1 + \gamma_2 Z_2 - 2/9X > U_M^0)$
$Z_1 = \mathbb{1}(P > 0)$	$Z_1 = \mathbb{1}(P > 0)$	$Z_1 = \mathbb{1}(P > 0)$
$Z_2 = \mathbb{1}\{Q > \mathbb{1}\{X > 5\}\}$	$Z_2 = \mathbb{1}\{Q > \mathbb{1}\{X > 5\}\}$	$Z_2 = \mathbb{1}\{Q > \mathbb{1}\{X > 5\}\}$
$D = Z_1$	$D = 1$	$D = 0$
$X = \max\{W^C   0 \leq W^C \leq X \leq 10\}$	$X = \max\{W^{AT}   0 \leq W^{AT} \leq X \leq 10\}$	$X = \max\{W^{NT}   0 \leq W^{NT} \leq X \leq 10\}$
$W^C = W + \mathbb{1}\{\mathcal{U}(0,1) > 0.85\}$	$W^{AT} = W - \mathbb{1}\{\mathcal{U}(0,1) > 0.85\}$	$W^{NT} = W$
$W \sim \mathcal{N}(8, 3)$	$W \sim \mathcal{N}(8, 3)$	$W \sim \mathcal{N}(8, 3)$
$P \sim \mathcal{N}(0, 1)$	$P \sim \mathcal{N}(0, 1)$	$P \sim \mathcal{N}(0, 1)$
$Q \sim \mathcal{N}(0, 5.08)$	$Q \sim \mathcal{N}(0, 5.08)$	$Q \sim \mathcal{N}(0, 5.08)$
$\gamma_1 \sim  \mathcal{N}(0, 1) $	$\gamma_1 \sim  \mathcal{N}(0, 1) $	$\gamma_1 \sim  \mathcal{N}(0, 1) $
$\gamma_2 = \gamma_1 + \eta + 0.2$	$\gamma_2 = \gamma_1 + \eta + 0.2$	$\gamma_2 = \gamma_1 + \eta + 0.2$
$\eta \sim  \mathcal{N}(0, 1) $	$\eta \sim  \mathcal{N}(0, 1) $	$\eta \sim  \mathcal{N}(0, 1) $
$N = 20,000$	$N = 15,000$	$N = 15,000$

The distribution of the observed confounder  $X$  is discrete, ranging between 0 and 10, and depends on the unobserved  $Z_1$  types. In total,  $X$  affects all observables. Note that also  $Z_1$  and  $D$  implicitly correlate with  $X$ , because the complier types have different  $X$  values. The unobserved components that affect the potential outcomes of  $Y$  and  $M$  have a common correlation structure between compliers as well as always and never takers of  $D$ . Their distribution reads:

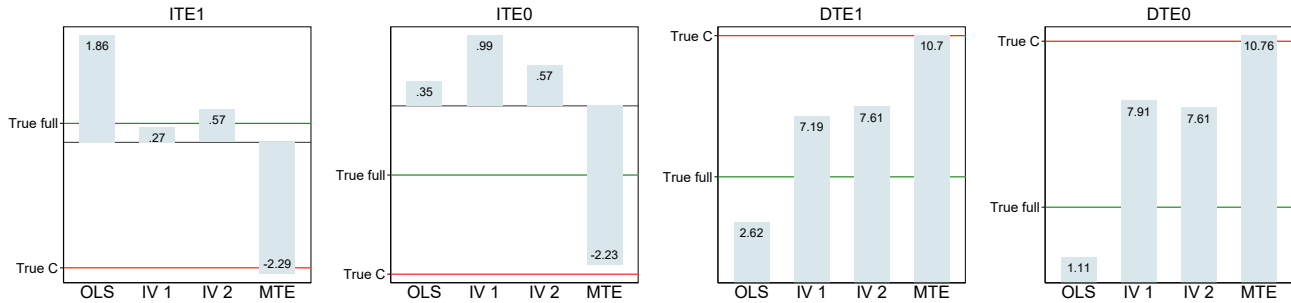
$$\begin{bmatrix} U^{11} \\ U^{10} \\ U^{01} \\ U^{00} \\ U_M^1 \\ U_M^0 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 1 \end{bmatrix} \right)$$

As before,  $M = M^1 D + M^0 (1 - D)$  and  $Y = D M Y^{11} + D (1 - M) Y^{10} + (1 - D) M Y^{01} + (1 - D) (1 - M) Y^{00}$ .

Results are shown in Figure B3. OLS and IV 2 are still far away from the true parameters. Now, however, in this more heterogenous DGP, classic IV with an interaction of  $D$  and  $M$

fails. *MTE* is now the only method that produces unbiased estimates of direct and indirect effects for compliers in this DGP. This DGP is complex in its type-specific heterogeneity, but it is not unlikely that this is a common feature of real-world data.

Figure B3: Results of DGP 3



Notes: Own calculations. This Figure plots the average results of DGP 3 after replicating the respective estimation procedures 200 times. Each panel refers to a specific indirect or direct mediation effect. The green and red horizontal lines depict the average true effect for the whole sample and the  $Z_1$  compliers only (our target parameter), respectively. OLS and IV1 refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by OLS and two-stage least squares (2SLS), respectively. IV2 reports the implied effect when no interaction term is used in the 2SLS outcome regression. MTE refers to the results of the estimation procedure derived in the paper. Full results, including RMSE, are reported in Table S1.

# Early- and Later-life Stimulation: How Retirement Shapes the Effect of Education on Old-age Cognitive Abilities

— For Online Publication: Supplementary Materials —

Hendrik Schmitz

RWI,  
Paderborn University

Matthias Westphal

University of Hagen,  
RWI

February 2025

---

Hendrik Schmitz: RWI – Leibniz-Institut für Wirtschaftsforschung and Paderborn University, Warburger Strasse 100, 33098 Paderborn, Germany, Tel.: +49 5251 603213, E-mail: [hendrik.schmitz@rwi-essen.de](mailto:hendrik.schmitz@rwi-essen.de).

Matthias Westphal: FernUniversität Hagen, Fakultät für Wirtschaftswissenschaft, Universitätsstraße 11, 58097 Hagen, Germany, Tel.: +49 2331 987-2633, E-mail: [matthias.westphal@fernuni-hagen.de](mailto:matthias.westphal@fernuni-hagen.de).

## Supplementary Materials A: Retirement rules

If not further mentioned, early retirement eligibility (ERA) criteria are mainly based on [Celidoni et al. \(2017\)](#) and information from the website of the United States Social Security Administration (<https://www.ssa.gov/policy/docs/progdsc/ssptw/>, accessed in August 2022) as well as *Pensions at a Glance* by the OECD and MISSOC (Mutual Information System on Social Protection Comparative Tables Database. <http://missoc.org>).

### Austria

**Men:** ERA is 60 for birth cohorts until 1940 (September). It was stepwise increased to 65 until the birth cohort 1952 (September).

**Women:** ERA is 55 for birth cohorts until 1945 (September). It increased stepwise to 60 until the birth cohort 1957 (September).

### Czech Republic

**Men:** 57 for birth cohorts until 1952 and 60 for older cohorts. Number of contribution years increased from 25 to 35 between cohorts 1952 and 1959.

**Women:** ERA depends on the birth year and the number of children. It is 50 for birth cohorts until 1957 with five children and 54 for birth cohorts until 1954 with 0. For later birth cohorts, ERA is stepwise increased to 60 (reached for birth cohort 1959 with up to two children and a bit later for more than two children).

**England:** Early retirement age is the same as official retirement age.

**Men:** 65 for birth cohorts until 1953, 65 and 10 months for birth cohort 1954 and 66 for birth cohorts 1955 and younger.

**Women:** 60 for birth cohorts until 1949. Gradually, it increased to 66 for birth cohorts between 1950 and 1954. 66 for birth cohorts 1955 and younger.

### France

**Men and Women:** ERA is 60 for birth cohorts until 1951 (June) and 62 for those born later.

### Germany

**Men:** ERA is 63 with at least 15 contribution years.

**Women:** ERA is 60 until birth cohort 1951 and 63 thereafter for those with at least 15 contribution years.

### Greece

**Men:** ERA is 62 with at least 15 contribution years.

**Women:** For women who started working before 1993: ERA is 55 with 15 contribution years. ERA is 50 for women with underage children and 18 contribution years. For women who started working since 1993: ERA is 60 with 15 contribution years. ERA is 50 for women with underage children and 20 contribution years.

### **Italy**

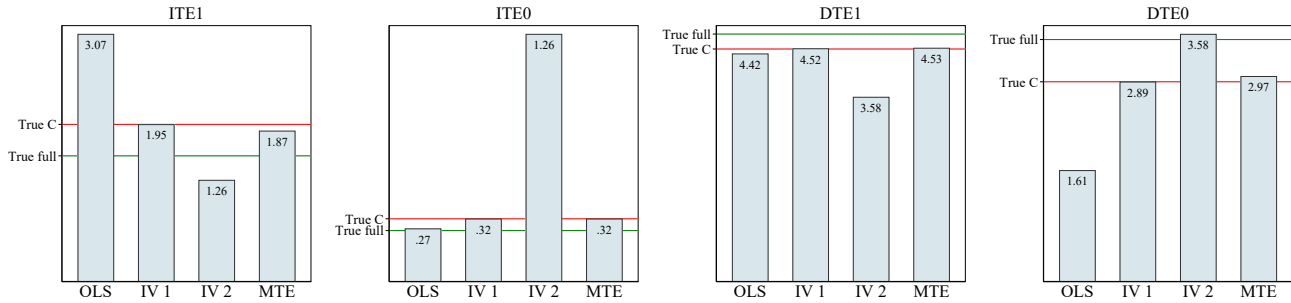
*Men and Women:* Between 1996 and 2012, stepwise increase from 52 (56 for self-employed) to 62. As of 2022: 64.

### **Spain**

*Men and women:* Until birth cohort 1947: 63. Stepwise increase until birth cohort 1960 to 65.

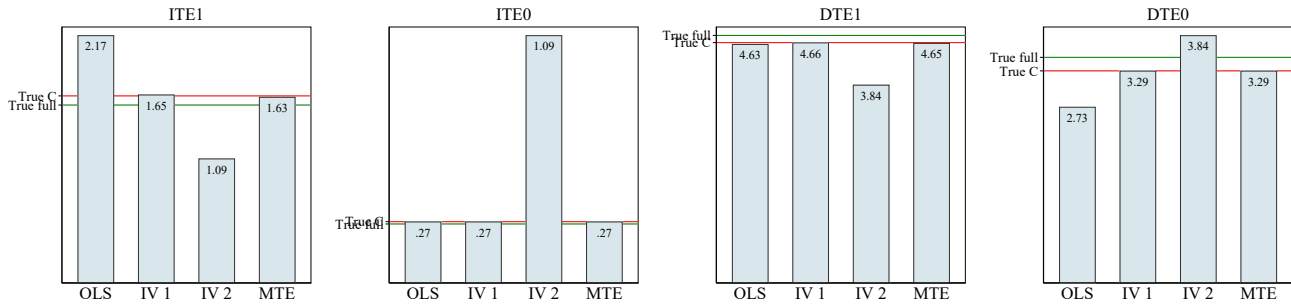
## Supplementary Materials B: More simulation results

Figure S1: Results of DGP 1'



Notes: Own calculations. This figure plots the average results of DGP 1 after replicating the estimation of the respective procedures 200 times. Each panel refers to a specific indirect or direct mediation effect. The green and red horizontal lines depict the average true effect for the whole sample and the  $Z_1$  compliers only (our target parameter), respectively. OLS and IV1 refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by OLS and two-stage least squares (2SLS), respectively. IV2 reports the implied effect when no interaction term is used in the 2SLS outcome regression. MTE refers to the results of the estimation procedure derived in the paper. Full results, including RMSE, are reported in Table S1.

Figure S2: Results of DGP 1''



Notes: Own calculations. This figure plots the average DGP 1'' results after replicating the respective estimation procedures 200 times. Each panel refers to a specific indirect or direct mediation effect. The green and red horizontal lines depict the average true effect for the whole sample and the  $Z_1$  compliers only (our target parameter), respectively. OLS and IV1 refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by OLS and two-stage least squares (2SLS), respectively. IV2 reports the implied effect when no interaction term is used in the 2SLS outcome regression. MTE refers to the results of the estimation procedure derived in the paper. Full results, including RMSE, are reported in Table S1.

Table S1 reports the results using 200 simulation rounds with 50,000 observations per round. The first two lines in each of the five blocks (denoted DGP 1, DGP 1', DGP 1'', DGP 2, DGP 3) report the (unconditional) average total, indirect and direct effects, both for the full sample as well as for the subsample of compliers to instrument  $Z_1$ . We aim to estimate the effects for the compliers and compare the average estimated effects to the average true effects.

Table S1: Simulation results

	TTE		ITE(1)		ITE(0)		DTE(1)		DTE(0)	
	$\widehat{E}[\cdot]$	RMSE	$\widehat{E}[\cdot]$	RMSE	$\widehat{E}[\cdot]$	RMSE	$\widehat{E}[\cdot]$	RMSE	$\widehat{E}[\cdot]$	RMSE
<i>DGP 1</i>										
True (full population)	5.06		1.56		.26		4.8		3.5	
True (compliers)	4.93		1.65		.27		4.66		3.29	
OLS	4.9	.75	2.17	9.08	.27	.18	4.63	.66	2.73	9.74
IV 1	4.94	.85	1.65	1.11	.27	.34	4.66	.81	3.29	.84
IV 2	4.94	.85	1.09	9.58	1.09	14.22	3.84	14.16	3.84	9.64
MTE	4.92	1.58	1.63	1.22	.27	.4	4.65	1.58	3.29	1.79
<i>DGP 1'</i>										
True (full population)	5.06		1.56		.26		4.8		3.5	
True (compliers)	4.84		1.95		.32		4.51		2.89	
OLS	4.69	2.65	3.07	19.51	.27	1	4.42	1.67	1.61	22.13
IV 1	4.84	.71	1.95	1.09	.32	.33	4.52	.69	2.89	.85
IV 2	4.84	.71	1.26	11.85	1.26	16.3	3.58	16.25	3.58	11.89
MTE	4.84	2.95	1.87	1.86	.32	.48	4.53	2.79	2.97	3.41
<i>DGP 1''</i>										
True (full population)	5.06		1.56		.26		4.8		3.5	
True (compliers)	4.93		1.64		.27		4.66		3.29	
OLS	4.89	.77	2.17	9.07	.27	.18	4.63	.66	2.73	9.73
IV 1	4.93	.89	1.65	1.14	.27	.34	4.66	.79	3.29	.74
IV 2	4.93	.89	1.09	9.59	1.09	14.19	3.84	14.15	3.84	9.61
MTE	4.92	1.64	1.63	1.26	.27	.38	4.65	1.58	3.29	1.61
<i>DGP 2</i>										
True (full population)	.98		.81		.57		.41		.16	
True (compliers)	1.44		.62		1.87		-.42		.82	
OLS	1.38	2.43	1.35	12.75	.14	29.93	1.24	28.89	.03	13.87
IV 1	1.44	4.19	.26	6.26	1.53	7.52	-.09	7.51	1.18	7.28
IV 2	1.44	4.19	.89	5.41	.89	17.13	.55	17.29	.55	6.13
MTE	1.48	4.67	.63	1.91	1.9	5.81	-.42	3.83	.85	3.42
<i>DGP 3</i>										
True (full population)	3.61		0.33		-0.97		4.58		3.29	
True (compliers)	8.32		-2.18		-2.36		10.68		10.50	
OLS	2.97	92.64	1.86	70.06	0.35	46.95	2.62	139.58	1.11	162.69
IV 1	8.16	4.26	0.27	42.50	0.99	58.16	7.19	60.66	7.91	45.08
IV 2	8.16	4.26	0.57	47.76	0.57	50.87	7.61	53.41	7.61	50.30
MTE	8.47	4.45	-2.29	3.54	2.23	4.02	10.70	4.39	10.76	6.39

Note: Simulation results on five different data generating processes (denoted DGP 1, DGP 1', DGP 1'', DGP 2, DGP 3) described in the paper using 200 rounds of simulation with 50,000 observations per round. The columns refer to the treatment parameters TTE, ITE(1), ITE(0), DTE(1), and DTE(0) that we evaluate. For every data-generating process, the first two lines refer to the true effects both for the full sample as well as for the subsample of compliers to instrument  $Z_1$ . The subsequent four lines refer to a separate estimator for the true effects. OLS and IV1 refer to implied mediation effects when  $Y = \delta_0 + \delta_D D + \delta_M M + \delta_{DM} DM + \delta_X X + \epsilon$  and  $M = \gamma_0 + \gamma_D D + \gamma_X X + u$  are estimated by OLS and two-stage least squares (2SLS), respectively. IV2 reports the implied effect when no interaction term is used in the 2SLS outcome regression. For all these estimators, estimate the TTE by an OLS or 2SLS regression of  $Y$  on  $D$ . Finally, the MTE estimator refers to the estimator derived in the paper. These estimators are evaluated by the point estimate ( $\widehat{E}[\cdot]$ ) and the root mean squared error (RMSE), which are presented in the columns. We aim to estimate the effects for the compliers and compare the average estimated effects to the average true effects.



## References

Celidoni, M., Bianco, C. D., and Weber, G. (2017). Retirement and cognitive decline. a longitudinal analysis using share data. *Journal of Health Economics*, 56:113–125.