Four Essays in the Economics of Education

Inaugural-Dissertation zur Erlangung des akademischen
Grades eines Doktors der Wirtschaftswissenschaft

doctor rerum politicarum (Dr. rer. pol.)

Fachbereich Wirtschaftswissenschaft

Freie Universität Berlin

vorgelegt von

Merlin Penny

Berlin, 2019

Erstgutachter: Prof. Dr. Viktor Steiner, Freie Universität Berlin

Zweitgutachter: Prof. Theocharis Grigoriadis, C.Sc. Ph.D., Freie Universität Berlin

Tag der Disputation: 11.07.2019

To my family and friends

爽歪歪

Acknowledgements

I would like to acknowledge the many individuals who have supported me throughout the last years. It is due to their efforts that I am able to complete this dissertation. I am very grateful for the support of my supervisor Viktor Steiner and very much appreciate being part of his seminar, having fruitful discussions and being provided with great freedom in developing my research ideas. I would also like to thank my second supervisor Theocharis Grigoriadis for his support.

I acknowledge financial support from the *Deutsche Forschungsgemeinschaft* and being part of the *Priority Programme 1646 Education as a Lifelong Process* which gave me the chance to participate in many interdisciplinary conferences. I am also honored to have been selected by the *Taiwan Fellowship* program and want to express my gratitude to Ming-Jen Lin and Jin-Tan Liu for the hospitality at National Taiwan University.

A large part of this dissertation was written as a researcher at the Department of Economics of the Free University of Berlin. I am grateful for the support of the faculty members, in particular Katja Görlitz, Ronnie Schöb, Dieter Nautz, Clemens Hetschko and Cordula Arlinghaus. I also thank Marcus Tamm at the RWI for good collaboration.

I am very grateful to Lukas Klingholz for the fantastic atmosphere in our shared office. I also thank fellow PhD candidates for their companionship, in particular Dominik Hügle, Astrid Harnack, Benjamin Fischer, Ksenia Anisimova, Tobias Wolf, Daniel Nachtigall, Marten von Werder, Maximilian Stockhausen, Maximilian Wenzel, Patrick Nüß and Tsu En Wang. Last but not least, I thank my family and many friends for their continuous support. My parents Sarah and Matthias Schairer-Penny, my siblings Myvanwy, Vasilii Pablo, Tallulah, Thelonious and Morgan and my beloved grandfather Alan Penny always believed in me. For that I am grateful.

Berlin, May 2019
Merlin Penny

Contents

Vorveröffentlichungen und Ko-Autorenschaften

Görlitz, K., M. Penny and M. Tamm (2019): The Long-Term Effect of Age at School Entry on Competencies in Adulthood. *Ruhr Economic Papers No. 792*, RWI.

Görlitz, K., M. Penny and M. Tamm (2019): The Long-Term Effect of Age at School Entry on Competencies in Adulthood. *DIW Discussion Papers No. 1789*, DIW Berlin.

Görlitz, K., M. Penny and M. Tamm (2019): The Long-Term Effect of Age at School Entry on Competencies in Adulthood. *Discussion papers / School of Business & Economics No. 2*, FU Berlin.

- **Chapter 1**. Merlin Penny:
  "Education and Work-Life Earnings"
- **Chapter 2**. Merlin Penny:
  "Grades in Primary School and Educational Opportunity - An Empirical Analysis for Germany"
- **Chapter 3**. Katja Görlitz, Merlin Penny and Marcus Tamm:
  "The Long-term Effect of Age at School Entry on Competencies in Adulthood"
- **Chapter 4**. Patrick Nüß and Merlin Penny:
  "Now You See Me! Ethnic Discrimination in the Market for Apprenticeships"

Rechtliche Erklärung

Hiermit erkläre ich, dass ich für diese Dissertation folgende Hilfsmittel und Hilfen verwendet habe und auf dieser Grundlage die Arbeit selbstständig verfasst habe. Die Arbeit wurde in keinem anderen früheren Promotionsverfahren vorgelegt, angenommen oder beurteilt.

- Stata
- Microsoft Office
- Google sheets
- LATEX
- Mendeley

Berlin, May 2019

Merlin Penny

RECHTLICHE ERKLÄRUNG

English Summary

Chapter 1 (Merlin Penny: *Education and Work-Life Earnings*) deals with the lifetime perspective of education and earnings whereof little is known since a large part of existing literature is concerned with estimating the returns to education. Therefore the purpose of this chapter is to calculate and analyze how age-earnings profiles and work-life earnings differ for educational groups. Work-life earnings describe the accumulated earnings from labor market entry to retirement. The main contributions are using longitudinal data with precise educational information to calculate work-life earnings and introducing a measure of work-life earnings that includes periods when individuals have no earnings due to not being employed. Foregone earnings during education are also calculated. A German dataset is used, i.e. survey data from the adult cohort of the National Educational Panel Study (NEPS). The survey data is linked to social security records (NEPS-SC6-ADIAB) to attain data on employment spells and earnings subject to the social insurance contribution. This data covers individuals that are born between 1944 and 1989 while the administrative data ranges from 1975 to 2011. The empirical proceeding is similar to related literature, i.e. wages are estimated that are censored due to being above the social insurance contribution limit. The selectivity of non-consent to record linkage is also investigated. Then the gross daily wages and days in employment in a year are used to create age-earnings profiles and calculate the present value of work-life earnings at age 18. The findings are that, on average, high skilled men earn 2.3 million EUR over the course of their working life. This is about 0.7 million EUR more than a medium skilled individual earns (1.7 million EUR). Foregone earnings of high skilled men during education as a fraction of work-life earnings are 3.9 percent. High skilled women earn, on average, only 0.3 million EUR more than medium skilled women. The results show that including periods where individuals are not employed

makes a substantial difference in work-life earnings. At the same time foregone earnings during education can also add up to a large fraction of work-life earnings.

Chapter 2 (Merlin Penny: *Grades in Primary School and Educational Opportunity - An Empirical Analysis for Germany*) investigates whether school reports should include grades or verbal assessments in the first years of primary school. A large part of existing pedagogical literature is concerned with this question. Very little is known, however, about the effects of actual reforms that led to later grading on education opportunity. In theory several channels are possible through which verbal reports could increase the likelihood of children with low parental education to transition to upper secondary school. The feedback parents receive in a verbal assessment is more extensive. Since cognitive costs of interpreting a school report are likely to differ for parents with different educational background this could mean that verbal assessments could also change educational opportunity. However, verbal reports could also be seen as a less precise signal concerning the cognitive abilities of a child. Other possible channels could be that teachers use verbal reports to strategically give distorted feedback to boost the confidence of students who are less confident due to their socio-economic background. The research question for this project is therefore: what effect does later grading in primary school have on educational transitions and educational achievement for students with different parental background. A reform is analyzed for Germany which meant federal states replaced grades in school reports by verbal assessments in the first classes of primary school starting in the 1970s. The data comes from the National Educational Panel Study (NEPS) for Germany as well as law gazettes of federal states. The NEPS includes 17,140 individuals born between 1944 and 1989 with information about educational transitions and family background. Law gazettes offer insight into the exact time of introduction of verbal reports. A difference-in-difference estimator is employed for the empirical strategy. This allows using variation given by the different timing of the implementation of the reform in federal states to identify a causal effect. Although results show that the coefficients are not significant the estimates provide some evidence for an effect of later grading on the transition to upper secondary school. Findings are robust to the inclusion of institutional rules concerning the transition to upper secondary school.

Chapter 3 (Katja Görlitz, Merlin Penny and Marcus Tamm: *The Long-term Effect of Age at School Entry on Competencies in Adulthood*) studies

whether school starting age (SSA) differentials continue into adulthood or fade away after leaving school. The previous literature has shown that children who enter school at a more advanced age outperform their younger classmates on competency tests taken between kindergarten and Grade 10. This chapter thereby contributes to this literature but also complements the literature investigating the long-run effects of SSA on individuals' wages and employment. The adult competencies are measured in comprehensive tests as part of a representative survey of individuals between 23 and 71. The identification of the effect of SSA on competencies relies on an instrumental variable strategy that exploits the state- and year-specific rules given by the cut-off dates. The empirical model controls for a full set of month-of-birth dummies. The model can also account for "age-at-test effects". This chapter contributes to the literature by additionally disentangling relative from absolute age effects, which is important from a policy perspective. Relative age measures the age difference compared to the ages of the other students within the cohort. Absolute age refers to the age (and, thus, maturity) when starting school. Results show that the impact of SSA on math and text comprehension measured in adulthood are considerably smaller than what the literature has shown for children in school. Further, both estimates are statistically insignificant. In contrast, the effect of SSA on receptive vocabulary is sizable in adulthood, with a one-year-higher SSA increasing competency by around a third of a standard deviation. These findings survive several tests of robustness. When disentangling the effect of SSA into an absolute and a relative age effect, we find that receptive vocabulary is affected solely by absolute age. Policies that shift the cut-off to an earlier date, making some children older by one year at school entry, should lead to an improvement in the average receptive vocabulary competencies even in adulthood.

Chapter 4 (Patrick Nüß and Merlin Penny: *Now You See Me! Ethnic Discrimination in the Market for Apprenticeships*) is about discrimination in the labor market for apprenticeships. Since individuals without German citizenship have only half the apprenticeship entry rate compared to those with German citizenship it is important to understand what role discrimination plays. In this chapter a correspondence experiment is used to study ethnic discrimination in the hiring market for apprenticeships in Germany. We send fictitious applications with German and Arabic-sounding names to real vacancies, augmenting applications with a short motivational video. These video introductions by applicants serve as a simple way of extending

employer insight into the personality and motivation of applicants and therefore solving the problem of discrimination in the hiring market for apprenticeships. This is done with the intention of decreasing potential statistical discrimination. Previous literature provide evidence for statistical, taste-based and attention discrimination in German-speaking labor markets. We find that applications with German-sounding names are on average twice as likely to receive an invitation for a job interview. Providing a video in the application that gives employers more insight into the personality and motivation of applicants increases invitations to job interviews for Arabic-sounding names. However, since these conditional effects are insignificant we interpret the results as weak evidence of videos increasing invitations for Arabic-sounding names. Furthermore, despite the fact that videos only have observable effects on Arabic-sounding names, the videos of German-sounding names are viewed 37% more often compared to the Arabic-sounding names, which is in support of attention discrimination.

German Summary

Die vier Kapitel dieser Dissertation sind voneinander unabhängig und beschäftigen sich mit verschiedenen Themen der ökonomischen Bildungsforschung. Das Vorwort und die jeweiligen Einleitungen der einzelnen Kapitel motivieren die unterschiedlichen Forschungsfragen und setzen sie in einen Zusammenhang mit der Literatur der Bildungsforschung.

Kapitel 1 (Merlin Penny: *Education and Work-Life Earnings*) handelt von der Lebensperspektive von Einkommen und Bildung worüber vergleichsweise wenig bekannt ist da ein Großteil bestehender Literatur sich mit dem Schätzen von Bildungsrenditen beschäftigt. Deshalb soll die Forschungsfrage beantwortet werden wie Alters-Verdienst-Profile und Lebenseinkommen sich nach Bildung unterscheiden. Lebenseinkommen beschreiben dabei die akkumulierten Einkommen vom Arbeitsmarkteintritt bis zum Renteneintritt. Der Beitrag des Kapitels liegt darin Längsschnitts Daten sowie zuverlässige Bildungsdaten zu verwenden und Perioden ohne Erwerbstätigkeit bei der Lebenseinkommensberechnung zu berücksichtigen. Neben der Berechnung von Lebenseinkommen werden auch entgangenen Einnahmen während eines Studiums berechnet. Als Datengrundlage dient die Erwachsenenkohorte des Nationalen Bildungspanels für Deutschland (NEPS) verknüpft mit administrativen Daten über sozialversicherungspflichtige Beschäftigungszeiten und deren Erwerbsbiographie (NEPS-SC6-ADIAB). Diese Daten umfassen Individuen welche zwischen 1944 und 1989 geboren sind während die administrativen Daten von 1975 bis 2011 reichen. Die Methodische Vorgehensweise orientiert sich an bestehender Literatur und beinhaltet das Schätzen von Löhnen welche über der Beitragsbemessungsgrenze der Sozialversicherung liegen und das Überprüfen einer möglichen Selektion bei der Zustimmung von Individuen zu der Verknüpfung der Daten aus der Sozialversicherung mit den Befragungsdaten. Darauf folgend werden

aus Bruttotageslöhnen und den Tagen in Beschäftigung in einem Altersjahr Alters-Verdienst-Profile erstellt und Barwerte von Lebenseinkommen zum Alter 18 berechnet. Ergebnisse zeigen, dass ein männlicher Universitätsabsolvent im Durchschnitt mit 2,3 Mio. EUR über das Erwerbsleben 0,7 Mio. EUR mehr verdient als eine Person die eine Lehre durchläuft (1,7 Mio. EUR). Die entgangenen Einnahmen von hochgebildeten Männern während der Ausbildung betragen etwa 3,9 Prozent vom Lebenseinkommen. Hochgebildete Frauen verdienen im Leben im Durchschnitt nur 0,3 Mio. EUR mehr als Mittelgebildete Frauen. Die Ergebnisse zeigen, dass es einen wesentlichen Unterschied macht welches Einkommenskonzept verwendet wird, das heißt ob lediglich Vollzeit-Beschäftigungszeiten betrachtet werden oder ob auch Perioden wo Individuen kein Einkommen haben hinzugefügt werden. Gleichzeitig können entgangene Einnahmen während eines Studiums einen Großteil der Lebenseinkommen ausmachen.

Kapitel 2 (Merlin Penny: *Grades in Primary School and Educational Opportunity - An Empirical Analysis for Germany*) beschäftigt sich mit der Forschungsfrage ob Schulzeugnisse in den ersten Jahren der Grundschule Noten oder Verbalbeurteilungen enthalten sollten. Dieser Frage widmet sich bereits ein Großteil bestehender pädagogischer Literatur, jedoch ist so gut wie nichts bekannt über die Effekte einer Schulreform welche zu einem späteren Zeitpunkt der Notengebung führt auf die Bildungschancen. In der Theorie sind mehrere Kanäle denkbar durch die Verbalbeurteilungen die Bildungschancen von Kindern mit niedriger elterlicher Bildung erhöhen. Generell bekommen Eltern durch verbale Beurteilungen ein umfassenderes Feedback über die kognitiven Fähigkeiten ihres Kindes. Wenn kognitive Kosten für Eltern mit niedriger Bildung bei Noten höher sind können Verbalbeurteilungen helfen Bildungschancen zu erhöhen. Auf der anderen Seite haben Lehrer durch Verbalbeurteilung die Möglichkeit gezieltes verzerrtes Feedback an Schüler mit niedrigem sozioökonomischem Hintergrund zu geben um das Selbstvertrauen zu stärken. Der Forschungsfrage welche sich dieses Kapitel widmet ist damit welche Effekte eine spätere Notengebung in der Grundschule auf Bildungsübergänge und Bildungsabschlüsse für Schüler mit unterschiedlicher elterlicher Bildung hat. Dazu wird eine Reform untersucht welche in Deutschland ab 1970 in den Bundesländern die Notengebung in Schulzeugnissen durch Verbalbeurteilungen in der ersten und in manchen Bundesländern auch in der zweiten Klasse ersetzt hat. Die Befragungsdaten kommen von der Erwachsenenkohorte des Nationalen Bildungspanels für Deutschland (NEPS) und umfassen 17.140 Individuen

welche zwischen 1944 und 1989 geboren wurden. Umfassende Information über Bildungsverläufe und Familienhintergrund sind darin vorhanden. Regelungen zu den Zeitpunkten der Einführung von Notenzeugnissen kommen aus den Gesetz- und Verordnungsblättern der Bundesländer. Die empirische Methodik beinhaltet einen Differenz von Differenzen Ansatz. Damit werden Variationen welche durch die unterschiedlichen Zeitpunkte der Einführung der Reform in den Bundesländern gegeben sind genutzt um kausale Effekte zu schätzen. Obwohl die Ergebnisse zeigen, dass die Koeffizienten nicht signifikant sind liefern die Schätzungen Belege für einen Effekt von einer späterer Notengebung auf den Übergang zur Sekundarschule. Die Ergebnisse sind robust wenn Informationen zu institutionellen Regelungen für den Übergang zur Sekundarstufe hinzugefügt werden.

Kapitel 3 (Katja Görlitz, Merlin Penny and Marcus Tamm: *The Long-term Effect of Age at School Entry on Competencies in Adulthood*) untersucht ob Unterschiede in Kompetenzen aufgrund des Alters bei der Einschulung bis ins Erwachsenenalter bestehen oder nach dem Schulabschluss bereits verschwinden. Frühere Literatur zeigt, dass Kinder, die bei der Einschulung älter sind als ihre Mitschüler, da sie deutlich vor dem Einschulungsstichtag geboren sind, besser als ihre jüngeren Mitschüler in Kompetenztests abschneiden, die erst kurz vor dem Einschulungsstichtag geboren sind. Diese Erkenntnis der bisherigen Literatur wurde bisher jedoch nur für Kompetenzen bis zur 10. Klassenstufe belegt. Die vorliegende Untersuchung betrachtet, ob sich der Einfluss des Alters bei der Einschulung auf Kompetenzen auch im Erwachsenenalter fortsetzt. Dazu wird eine Befragung aus Deutschland genutzt, die Kompetenztests in Mathematik und Deutsch umfasst. Die Kompetenztests werden in umfangreichen Tests welche Teil einer repräsentativen Befragung für Individuen zwischen 23 und 71 sind gemessen. Die Identifikation der Effekte greift auf Variation im Einschulungsstichtag zwischen Bundesländern und über die Zeit hinweg zurück. Dafür wird ein Instrumenten Variablen Schätzer angewandt während das Empirische Modell auch für Indikatoren für die Geburtsmonate kontrolliert. Des weiteren berücksichtigt das Modell auch Effekte für das Alter zum Zeitpunkt eines Tests und separiert relative von absoluten Alterseffekten. Letzteres ist relevant um Politikempfehlungen treffen zu können. Das relative Alter misst die Unterschiede im Vergleich zu dem Alter der anderen Schüler in der selben Kohorte. Im Gegensatz dazu bezieht sich das absolute Alter auf das Einschulungsalter (und damit Reife) beim Schulstart. Die Ergebnisse zeigen, dass

die Effekte des Einschulungsalters auf Kompetenzen in Mathe und im Textverständnis im Erwachsenenalter wesentlich geringer sind als jene welche die Literatur für Kinder im Schulalter findet. Des weiteren sind beide Schätzungen statistisch insignifikant. Im Gegensatz dazu ist der Effekt des Einschulungsalters auf rezeptiven Wortschatz im Erwachsenenalter ausgeprägt. Ein um ein Jahr höheres Einschulungsalter erhöht hier die Kompetenz um etwa ein Drittel einer Standardabweichung. Die Ergebnisse bestehen mehrere Robustheitstests. Beim Entflechten von absoluten und relativen Alterseffekten zeigt sich, dass die Effekte für den rezeptiven Wortschatz ausschließlich vom absoluten Alter getrieben sind. Politikmaßnahmen welche den Einschulungsstichtag zu einem früheren Datum verschieben und damit einige Kinder bei der Einschulung um ein Jahr älter machen sollten daher zu einer Verbesserung in den durchschnittlichen Kompetenzen für den rezeptiven Wortschatz im Erwachsenenalter führen.

Kapitel 4 (Patrick Nüß and Merlin Penny: *Now You See Me! Ethnic Discrimination in the Market for Apprenticeships*) untersucht Diskriminierung im Arbeitsmarkt für Lehrstellen. Da Individuen ohne deutsche Staatsangehörigkeit nur etwa eine halb so große Ausbildungseintrittsrate haben verglichen zu denen mit deutscher Staatsangehörigkeit ist es wichtig zu verstehen welche Rolle Diskriminierung spielt. In diesem Kapitel wird ein so genanntes *Correspondence Experiment* verwendet um ethnische Diskriminierung im Bewerbermarkt für Lehrstellen in Deutschland zu erforschen. Wir senden fiktive Bewerbungen mit Deutsch und Arabisch klingenden Namen zu echten Vakanzen und fügen Bewerbungen kurze Motivationsvideos hinzu. Diese Videovorstellungen der Bewerber dienen als ein einfacher Weg dem Arbeitgeber einen tieferen Einblick in die Persönlichkeit und Motivation von Bewerbern zu geben und somit zu der Lösung des Problems der Diskriminierung im Bewerbermarkt für Lehrstellen beizutragen. Dies wird mit der Absicht getan potentielle statistische Diskriminierung zu verringern. Frühere Literatur findet für deutschsprachige Arbeitsmärkte Belege für statistische Diskriminierung, Diskriminierung basierend auf einer persönlichen Präferenz eines Arbeitgebers (*taste-based*) sowie Diskriminierung beruhend auf der Aufmerksamkeit mit welcher ein Arbeitgeber eine Bewerbung sichtet. Die Ergebnisse zeigen, dass Bewerber mit Deutsch klingenden Namen im Durchschnitt eine doppelt so hohe Wahrscheinlichkeit haben zum Vorstellungsgespräch eingeladen zu werden. Das Beifügen eines Videos in der Bewerbung erhöht die Wahrscheinlichkeit für Individuen mit Arabisch klingenden Namen zum Vorstellungsgespräch eingeladen zu werden.

Da diese bedingten Effekte jedoch insignifikant sind interpretieren wir die Ergebnisse als schwache Evidenz dafür, dass Videos Einladungen für Arabisch klingende Namen erhöhen. Obwohl die Videos nur Effekte auf Einladungen für Arabisch klingende Namen haben werden Videos von Personen mit Deutsch klingenden Namen trotzdem 37% häufiger betrachtet was für eine Diskriminierung beruhend auf der Aufmerksamkeit spricht.

List of Figures

LIST OF FIGURES

List of Tables

Preface

This dissertation with the title *Four Essays in the Economics of Education* studies a broad spectrum of research questions. The topics covered range from the documentation of earnings over the working life to school grading, school starting age and discrimination in the labor market for apprenticeships. Therefore the chapters can in some way be seen as a reflection of the evolution in the economics of education where after many years of research on the monetary returns to education more and more research is devoted to cognitive and non-cognitive skills. At the same time behavioral and experimental economics have also started playing a more important role in educational research.

The relevance of estimating the returns to education becomes apparent when regarding the strong positive association of educational levels with labor market, health and social outcomes (Heckman et al., 2011). The theoretical basis for the monetary returns to education is laid by early work of Becker (1964), Schultz (1961) and Mincer (1958) on human capital theory where investments into human capital increase productivity and wages. The years of schooling are an essential part of this concept and have until now been the most used measure for human capital acquired. I contribute to this literature by documenting work-life earnings while also focusing on calculating foregone earnings which are opportunity costs of education and therefore important indirect costs. While the educational achievement has a strong association with earnings (Card, 1999) these again are positively associated to other long-term outcomes, e.g. life satisfaction (Kahneman and Deaton, 2010) and longer lifetime (Chetty et al., 2016). At the same time additional schooling causally increases the performance in intelligence tests in adulthood (Carlsson et al., 2014). This is relevant since cognitive skills are known to also explain labor market outcomes (Heckman et al., 2006) and suggests that lifetime earnings and cognitive skills can be seen as equally important labor market outcomes. I study both in separate chapters of this dissertation with different research questions.

PREFACE

The importance of studying family inputs in the economics of education becomes clear when considering that there is ample evidence for a strong association between parental background and educational paths in Germany (Bos et al., 2007; Dustmann, 2004). Even though some of this research indicates that early educational track choices can be revised (Dustmann et al., 2017) complementing research does suggest that later career choices remain highly selective with respect to parental background (Biewen and Tapalaga, 2017). In fact, family inputs are one important component in the education production where a student's achievement can be seen as a function of different inputs that also include school and peer inputs as well as individual abilities (Koch et al., 2015). I study parental background in the context of determining the optimal time to start grading in primary school and the effect of later grading on educational opportunity. However, family inputs also play a role in this dissertation for instance for the school starting age where parents can delay the school entry of their child by a year if they believe that this is an optimal decision.

The strong relationship between parental background and educational career paths can be reinforced by expectations teachers tend to have for the performance of students with different parental background (Figlio, 2005). Lower grades given to pupils with foreign origin (Sprietsma, 2013) show discrimination in this area of research. The question I study which type of feedback (grades or verbal reports) helps disadvantaged families the most to improve the educational achievement of their children is therefore related to this. Discrimination on the basis of ethnicity and socio-economic status is not only found in form of grading and teacher recommendations but also in many other areas of life. One of these is the labor market hiring process where for instance youths with a migration background applying for an apprenticeship are much less likely to be invited to a job interview than those without migration background (Beicht, 2016). Avoiding initial small failures in school and the apprenticeship labor market caused by discrimination becomes even more important considering that these can lead to further academic or even vocational failures through self-fulfilling prophecies (Glover et al., 2017). Therefore, the last chapter of this dissertation studies whether video applications can reduce discrimination in the apprenticeship labor market.

The data used in this work are from the National Educational Panel Study (NEPS) for Germany as well as data collected in a field experiment for Germany. Therefore the

German education system and its development over time is of special relevance. Education in Germany is characterized by the educational autonomy of the federal states which was granted for the West German states in 1949. In the following decades up until now a multitude of educational reforms were introduced at different points in time. Some of the reforms during this time include grading reforms as well as cutoff dates that effect the school starting age. These are addressed in Chapter 2 and Chapter 3. Both evaluations rely methodologically on exploiting the variation of the time of implementation of the reforms between states. The adult cohort of the NEPS includes the educational biographies of individuals born as far back as 1944 and therefore offers an optimal data set for the analysis and a basis for Chapter 1 where the work-life education-earnings for far-ranging cohorts are calculated. By evaluating reforms that took place in Germany I contribute to the human capital literature that does not use US data. Since the education system in Germany differs to that in other countries, e.g. students do not have high direct costs resulting from education like in the US, it is necessary to increase the number of studies that look at different institutional settings.

The chapters in this dissertation differ not only in the topics covered but also in the empirical methods used depending on the research question, data availability and structure. The empirical methods applied range from descriptive evidence (Chapter 1) to the more causal difference-in-differences (Chapter 2) and instrumental variable analysis (Chapter 3) and finally a correspondence experiment (Chapter 4). Table 1 gives an overview of the research questions, data, methods and results of all four chapters. Identifying causal treatment effects in educational research is necessary to be able to make policy recommendations. For instance, for policy makers to decide whether to increase or decrease school starting age they can not simply rely on a simple estimation of school starting age and years of education or other long-term outcomes. Since the school starting age is endogenous and dependent on a cutoff date it is necessary to either rely on a instrumental variables or a regression discontinuity design. The reason for this is that children who belong to the younger children of a cohort are not directly comparable to their classmates who are up to one year older when starting school.

Researchers often use natural experiments to identify causal treatment effects when they are not able to perform randomized controlled trials. For many research questions it is not possible to perform field experiments due to the simple fact that a researcher faces limited time and financial resources. At the same time studying school starting

Table 1: Chapter overview

Chapter 1	**Chapter 2**	**Chapter 3**	**Chapter 4**
Title			
Education and Work-Life Earnings	Grades in Primary School and Educational Opportunity - An Empirical Analysis for Germany	The Long-term Effect of Age at School Entry on Competencies in Adulthood	Now You See Me! Ethnic Discrimination in the Market for Apprenticeships
Main research question			
Documenting the lifetime perspective of education and earnings.	What is the effect of abolishing grades in primary school reports on educational opportunity?	Do effects of school starting age continue into adulthood?	Is there evidence of ethnic discrimination in the labor market for apprenticeships? Can motivational videos in applications reduce ethnic discrimination?
Methods			
Descriptive analysis	Difference-in-differences analysis	Instrumental variable analysis	Field experiment
Data			
National Educational Panel Study adult cohort linked to social security records (NEPS-SC6-ADIAB)	National Educational Panel Study (NEPS), starting cohort adults (SC6)	National Educational Panel Study (NEPS), starting cohort adults (SC6)	Correspondence experiment
Main findings			
The highest work-life earnings are found for high skilled men while a large gender work-life earnings gap is found. Foregone earnings can constitute a considerable fraction of work-life earnings. However, results strongly depend on the sample restrictions.	Results show that the coefficients are not significant, however, the estimates still provide some evidence for an effect of later grading on the transition to upper secondary school. Findings are robust to the inclusion of institutional rules concerning the transition to upper secondary school.	The effects of school starting age render insignificant for math competences and text comprehension in adulthood. The effect is sizable for receptive vocabulary.	Applications with German-sounding names receive twice as many callbacks as those with Arabic-sounding names. However, videos increase callbacks for Foreign-sounding names by over one third. Results can be seen as weak evidence that videos reduce statistical discrimination.
Own contribution			
100 %	100 %	33.3 % (in co-authorship with Katja Görlitz and Marcus Tamm)	50 % (in co-authorship with Patrick Nüß)

age or grading in school reports would require randomly assigning individuals to treatment and control groups and would therefore be ethically questionable. Studying ethic discrimination in the hiring market for apprenticeships on the other hand allows the use of a field experiment which identifies causal treatment effects. Such a correspondence experiment can also be conducted in a reasonable time frame since researchers are only interested in whether the applicant is invited to a job interview or not and are not estimating long-term effects.

This dissertation commences by studying one of the most prominent research questions in the economics of education, i.e. the general relationship between education and work-life earnings (Chapter 1). Documenting the lifetime perspective of education and earnings has in the past rarely been done due to data availability. Most studies in the past focus on using cross-sectional data and estimating the coefficient of the years of schooling in earnings functions. I present results of age-earnings profiles, work-life earnings and foregone earnings and make a distinction by using different samples for individuals who are employed throughout their career and those who are not.

The following Chapter 2 studies school grading and educational opportunity, topics which have during recent years become more and more important in educational research. I contribute to a growing amount of research in this field by analyzing the effect of abolishing grades in primary school reports on educational opportunity as measured by secondary school track choice for students with different parental education. The results are especially interesting for countries with early tracking systems and countries where educational achievement is highly dependent on parental background.

In Chapter 3 the causal effect of school starting age on adult competencies is studied. While there exists ample research on the effects of school starting age on in-school outcomes little is known about the long-term effects of age at school entry on competencies in adulthood. This chapter therefore contributes to a growing amount of educational research on cognitive skills and complements existing studies that analyze the effect of school starting age on earnings. While previous literature suggests that the effects of age at school entry get smaller as children grow older little is still known whether these effects remain important during adulthood. The competencies focused on are mathematical literacy, text comprehension and receptive vocabulary. At the same time the chapter also contributes to the literature by separating relative from absolute age effects.

Finally, Chapter 4 consists of a correspondence experiment that aims to find out whether video applications can diminish ethnic discrimination in the labor market for apprenticeships. In Germany, individuals without German citizenship have only half the apprenticeship entry rate compared to those with German citizenship (BMBF, 2018). This is explained with differences in occupational preferences, regional labor markets, limited proficiency in the German language and participation rate in internships. The chapter aims at finding out what role discrimination plays. Given the importance of the school-to-work transition for future labor market prospects, it is important to understand the relevance of discrimination and to search for solutions in the school-to-work transition. Therefore a field experiment is conducted by sending applications to vacancies for apprenticeship positions. The applications differ in the applicants ethnicity and the inclusion of a short motivational video. This allows employers to gain insight into the personality of applicants and therefore offers a solution to discrimination in the hiring market for apprenticeships. Previous experiments have found both statistical discrimination (Kaas and Manger, 2012) and taste-basted discrimination (Weichselbaumer, 2016a) in German-speaking labor markets. By including videos in applications it can be tested whether potential statistical discrimination can be decreased.

Chapter 1

Education and Work-Life Earnings

1.1 Introduction

Empirically, higher levels of education are generally associated with higher labor market, health and social outcomes (Heckman et al., 2011). The most prominent association, however, is that of the relationship between education and earnings (Card, 1999). Higher earnings have not only been found to improve individual evaluation of life (Kahneman and Deaton, 2010), but to lead to greater longevity (Chetty et al., 2016) while higher relative earnings increase well-being (Luttmer, 2005). High work-life earnings also have a positive impact on retirement benefits from social security contributions (Tamborini et al., 2009). Hence, the calculation of work-life earnings, which are defined as the cumulative yearly earnings from labor market entry until retirement, is of special interest since individuals desire to maximize their utility of lifetime earnings by choosing a certain level of education. The theory of investment in human capital which was introduced by Becker (1964) lays the theoretical basis for the calculation of work-life earnings. In his model, investment in education and training increases productivity and wages. At the same time this model and also the school-work model by Ben-Porath (1967) emphasize the importance of foregone earnings during training.

In the present study I document differences in age-earnings profiles and work-life earnings for educational groups. Data from the National Educational Panel Study (NEPS) are linked to longitudinal administrative social security records for West Germany. In this way reliable educational information from survey data and employment variables from administrative data can be used. The paper aims at contributing to the

literature on work-life earnings in respect to education in the following ways:

Firstly, the aim of this paper is to find out how work-life earnings and age-earnings profiles differ for educational groups. The calculation of work-life earnings includes finding out how large a fraction of work-life earnings is made up of the opportunity costs of education, i.e. foregone earnings of students. Many studies have come to the conclusion that special attention should be paid to earnings that university students forgo by not working (Ben-Porath, 1967; Becker, 1964; Schultz, 1960). However, foregone earnings have in the past mostly not been calculated separately because they are implicitly included in returns to education estimates (Heckman et al., 2006; Card, 1999).

Secondly, data limitations hinder the analysis of work-life earnings to a large extent. This study contributes to the literature through the use of panel data which allows the calculation of work-life earnings of several generations for men and women. In the past studies calculating work-life earnings have, for the most part, relied on cross-sectional pooled data due to far-ranging longitudinal data being unavailable (Schmillen and Stüber, 2014; Julian and Kominski, 2011; Carnevale et al., 2011; Day and Newburger, 2002). These studies assume that earnings of older and younger cohorts are similar when calculating work-life earnings. Other studies that calculate work-life earnings with panel data either analyze different countries, e.g. Tamborini et al. (2015) for the US, or focus on measures of inequality instead of educational levels (Bönke et al., 2015).

Thirdly, a contribution is made by introducing a holistic measure of earnings that is not restricted to certain subgroups of working individuals, e.g. full-time or full-year employed, and can therefore yield work-life earnings that are representative for the entire population. The measurement of earnings has been discussed by Card (1999) who illustrates that the return to schooling is almost always higher for weekly and annual earnings than it is for hourly wages because individuals with higher levels of schooling work more. He notes that about two-thirds of the returns to education calculated by using annual earnings data is due to the effect of education on hourly wage while the rest is due to hours per week and weeks per year. While research differs in the working time restrictions made for calculating work-life earnings, in most studies average earnings of full-time workers are examined (Schmillen and Stüber, 2014; Day and Newburger, 2002). This restriction neglects the fact that labor supply of individuals changes to a large extent during their working career for reasons such as unemployment, motherhood,

early retirement and other circumstances. Hence, these are normally higher than average work-life earnings for all workers, including part-time individuals and individuals that work only part of the year (Day and Newburger, 2002). Therefore, different sample restrictions are used here: The first sample is restricted to spells where individuals are in an employment. The second sample includes periods where individuals have no earnings due to voluntary or involuntary reasons. The distinction made here can answer questions regarding the impact of accounting for periods where individuals have zero earnings.

The chapter is structured as follows: Section 1.2 provides an overview of the relevant literature. Section 1.3 describes the data and methodology. Section 1.4 examines age-earnings profiles and work-life earnings. Section 1.5 concludes.

1.2 Relevant literature

1.2.1 Human capital theory

The human capital theory literature starts devoting more relevance to work-life earnings in the 1960s. Theoretical models developed by Becker (1964, 1962) also start to distinguish between direct costs and foregone earnings in human capital investment models for workers receiving on-the-job training. In his model individuals invest in their human capital by choosing an educational level in order to increase future productivity and wages. After comparing expected work-life earnings of these career decisions and also considering the direct and indirect costs during education an educational level is chosen.[1] Human capital theory also predicts that earnings increase with experience and additional investments in human capital are worth less at the end of the career (Mincer, 1974). Ben-Porath (1967) develops a schooling choice model where the human capital accumulation is modeled as an optimal path of investment in human capital over an individual's life cycle. The human capital production function thereby explains the growth of earnings with age. This model implies that late investments yield smaller benefits because of the shorter time span in which any gains of these investments can be reaped. This explains why a rational decision of allocating time to investments in education means investing at a young age (Ben-Porath, 1967).

[1] Taking into account preferences for leisure.

1.2.2 Work-life earnings and other measures

The empirical literature calculating work-life earnings by educational groups mostly uses cross-sectional data. While studies for the US find individuals with a bachelor degree earn on average between 2-3m USD during their career, a recent study for Germany finds work-life earnings of 2.3m EUR for individuals with a university degree (Schmillen and Stüber, 2014). Day and Newburger (2002) use data from the Current Population Survey and find that average earnings for individuals who work full-time and year-round are higher than average earnings of all workers including people who work part-time or for a part of the year. They also find a substantial gap between work-life earnings of men and women where women completing a bachelor degree earn about 1m USD less than men with the same level of education. Julian and Kominski (2011) use the American Community Survey to find similar results for more recent years (2006-08). The authors document the rise of the level of education in America in the last 70 years and strong differences in the earnings ratios for gender and ethnicity. Carnevale et al. (2011) confirm these results in an extensive study of work-life earnings by educational attainment. While these calculations are for full-time and full-year workers, the authors note that many workers, especially women[2], do not meet this criteria. Other studies from the US use lifecycle earnings models to estimate work-life earnings: Webber (2016) calculates expected work-life earnings for a median male person for college majors including explicit and implicit costs of education and finds that a four-year STEM (science, technology, engineering or math) degree with average college expenses has a value of about 2.1 million US Dollars. Examining data from the federal pension insurance register for Germany, Bönke et al. (2015) offer some insight into age-earnings profiles for educational groups and pooled cohorts for the years 1935-49 and find higher earnings growth for higher educated individuals throughout their career. However, the study does not present any estimates of work-life earnings. Tamborini et al. (2015) use data for the US that span over 26 years. They also include individuals without positive earnings and find large differences in work-life earnings especially for women. However, the interpretation of most of the mentioned studies using pooled cross-sectional profiles is very different to that of longitudinal profiles. The estimates from Table 1.1 give work-life earnings for educational groups and answer the question as to how much individuals

[2]Rose and Betts (2004) find that women in the US work on average 500 hours less than men per year.

Table 1.1: Reports of lifetime earnings

Study	Lifetime earnings	Age Range	Data	Years
Day and Newburger (2002)	2.3m USD[a]	25-64	CPS	1997-1999
Julian and Kominski (2011)	2.6/2.8m USD[b]	25-64	ACS	2006-2008
Carnevale et al. (2011)	2.2m USD[c]	25-64	ACS	2007-2009
Schmillen and Stüber (2014)	2.3m EUR[d]	25-64	SIAB	2008-2010
Tamborini et al. (2015)	2.4m USD[e]	20-69	SIPP-IRS	1982-2008

Notes: The lifetime earnings are those for holding a bachelor degree (a, b, c) or other university degree (d). a) Averages b) Medians for white males; lower value represents all workers, higher value represents full-time year round workers. c) Median full-time year round workers d) Averages of full-time workers. e) Median lifetime gross earnings for men Data: CPS: Current Population Survey; ACS: American Community Survey; SIAB: Stichprobe der Integrierten Arbeitsmarktbiografien (i.e. administrative data for Germany); SIPP-IRS: Survey of Income and Program Participation linked to Detailed Earnings Record.

with the same educational level are expected to earn in a hypothetical career. Thereby assuming that cross-sectional earnings are representative for future earnings. This does not depict changes in work-life earnings of birth cohorts. Interpreting these findings as representative for a population would require the assumption that work-life earnings do not differ between cohorts.[3] Another assumption is that there is no increase in productivity in the economy. Finally, it is assumed that individuals participate in the labor force without interruption and work full-time and full-year. The present study, similar to Tamborini et al. (2015), uses a less restrictive sample and includes individuals who are not as attached to the labor market.

Apart from calculating work-life or lifetime earnings there are other measures used in the literature. Since the "Mincerian" earnings function was introduced (Mincer, 1974), a very strong focus has been placed on estimating the coefficient of the years of schooling by using this method. While this method also implicitly accounts for foregone earnings[4] there have been different methods in the literature to calculate returns. Various types of returns and treatment effects of higher education have been estimated, which vary,

[3]For instance, a person who has annual earnings in the year 2010 at age 30 is assumed to have the same work-life earnings in 2030 as a person aged 50 in the year 2010.

[4]Schultz (1960) is one of the first to estimate foregone earnings of students in high school, college and university. He finds that foregone earnings are 43% of total costs in 1956 and are therefore as important as direct costs.

depending on which treatment effects one is interested in (Dickson and Harmon, 2011; Heckman et al., 2006). However, the "Mincerian" earnings function has been criticized for a number of reasons. Heckman et al. (2006) argued that, in recent decades, the assumption that log-earnings experience profiles are parallel across schooling levels has failed.[5] The function has also been criticized by Murphy and Welch (1990) who show that earnings growth changes to a larger extent early in a career than later in a career and therefore does not decline linearly and is not constant. The advantage of the method is that causal estimates can be achieved if a source of exogenous variation is found to identify wage returns.[6] While this has become the preferred empirical method when estimating the coefficient of the years of schooling in an earnings function, the "full-discounting" or "elaborate" method which was mainly used in the sixties is not in wide use anymore. It compares lifetime earnings streams of different educational levels by using individual age-earnings profiles (Psacharopoulos, 2014; Becker, 1964) and has the advantage that less assumptions are required. Since it is possible to construct these profiles here this method will also be used in the present study. However, the focus of this study remains providing descriptive evidence of age-earnings profiles and work-life earnings.

1.3 Data and Methodology

1.3.1 Data

The data originate from the adult cohort of the German National Educational Panel Study (NEPS), as described in Blossfeld et al. (2011) that are linked to social security records by the German Institute for Employment Research (IAB).[7] NEPS-SC6-ADIAB is unique, as educational survey data is combined with retrospective administrative information on earnings and employment over the working lives of individuals. The NEPS Adult Starting Cohort was started in 2007 and covers 17,140 individuals born

[5] Even though the predictions that log-earnings age profiles diverge with age across schooling levels and the variance of earnings over the life cycle has a U-shaped pattern do hold for recent data (Heckman et al., 2006).

[6] Some examples for natural experiments include month of birth in combination with compulsory schooling laws (Angrist and Krueger, 1991) or twins (Ashenfelter and Krueger, 1994; Leigh and Ryan, 2008). However, as Card and Lemieux (2001) note, true randomization can almost never be achieved completely and one should therefore interpret parameter estimates with caution.

[7] The structure of the administrative data in NEPS-SC6-ADIAB is described in Antoni et al. (2018).

between 1944 and 1989. Up to the year 2013 five NEPS waves are available which offer detailed information on educational biographies, family background and employment. From this data set the highest educational degrees are used for the analysis. The retrospective collection of individuals' educational information in the NEPS can be seen as reliable since there is a data revision module.[8]

The administrative data consists of social security records for the years 1975 to 2011. The data includes individuals who have one of the following statuses: employment subject to social security contributions (recorded from 1975 onwards), marginal part-time employment (recorded from 1999 onwards), receipt of benefits (recorded from 1975 onwards) or registered as a job seeker at the Federal Employment Agency (recorded from 2000 onwards). The administrative data does not include civil servants and self-employed workers. The dataset is largely comprised by the original data without having undergone anonymization processes.

A probabilistic record linkage is performed by the IAB, which also provides the final dataset. While the willingness of individuals to agree to be linked with the IAB data is at 88.77 percent in the NEPS, the IAB reports that 91.35 percent of these willing individuals were able to be matched by an exact, probabilistic or manual match with social security records. These are all workers and employees. Compared to other surveys, e.g. Sakshaug and Kreuter (2012), there is a high consent rate. Still, selectivity of non-consent to record linkage can be a source of bias if education and earnings differ for the groups that agree or disagree to linkage. Here, Table 1.A.2 shows the differences in means of characteristics. Applying a t-test reveals that both groups are similar in terms of birth year, gender, years of education and years of education of parents. However, the share of individuals who are married is slightly higher for individuals that agreed to the linkage. This group is also less often self-employed. Another way to test non-consent selectivity in a regression framework is to find an exclusion restriction and test for selection bias as suggested by Fertig and Görlitz (2013). The idea they propose is to find a measure that identifies selection and apply a Heckman selection model. The exclusion restriction applied here is the standard deviation of the interviewer specific months which can be used as a measure of interviewer engagement in the survey. The consideration behind this is that a low standard deviation is evidence of an interviewer

[8] After each interview a data revision is done and inconsistencies in the sequence of episodes are identified. The respondent is then asked to add missing events and dates (Skopek, 2013).

Table 1.2: Data restrictions

	Observations
Employee history (Employment spells)	363,781
Without interns and student assistants in firms (since 1997)	360,421
Without individuals below the SCC ceiling (since 1999)	318,142
Without part time older workers (since 2000)	315,183
Only employees subject to SCC and those in vocational training	313,776
Without year 2012 (missing establishment data)	302,414
Spells split on monthly level	2,484,249

Notes: Dataset NEPS-SC6-ADIAB. SCC: social security contribution.

knowing specific issues of the questionnaire better. Table 1.A.1 show the results. The coefficient of the exclusion restriction is statistically significant in the selection equation while a higher standard deviation of months conducting the interviews is negatively related to the individual's likelihood to give consent. Furthermore, ρ is insignificant indicating no statistically significant correlation between the unobservables of the wage and the selection equation.

Other data quality issues are censored wages in the social security data as well as some inconsistencies over time. The censoring is solved by imputing top coded wages. Similarly to the literature (Reichelt, 2015; Dustmann et al., 2009; Gartner, 2005) this is done by conducting a regression, predicting the right censored wages and adding an error term. A detailed explanation of this can be found in Section 1.A.1.

In 1984 one-time payments became subject to social security tax and are therefore only included in the administrative data starting from 1984. Steiner and Wagner (1998) note that that ignoring this can increase earnings inequality. However, in the dataset the structural break does not affect many individuals as is argued in Section 1.A.2.

The following data restrictions are done as depicted in Table 1.2: From the administrative data for the years 1975-2011, employees subject to social security contributions and those in vocational training are retained. The data is used only up to the year 2011 because establishment data which is indispensable in the regressions imputing censored wages is missing for 2012. The following groups are dropped from the sample because

Table 1.3: Samples

		Men	Women	Total
Individuals		5,156	5,101	10,266
Panel	*Panel structure*			
1. Employed (EMPL)	Unbalanced	104,111	86,957	191,068
2. Employed plus (EMPL+)	Balanced	143,596	141,988	285,584

Notes: West German men and women; Dataset NEPS-SC6-ADIAB. Panel-years refer to yearly earnings for years 1975-2011 and ages 18-64.

they are not available from 1975 onwards: individuals below the social security contribution ceiling (added in 1999), interns and student assistants in firms (since 1997) and part time older workers (since 2000). Only taking into account employees subject to social security contributions and those in vocational training up to the year 2011 (since establishment data is only available until 2011) leaves 302,414 employment spells which relate to 2,484,249 employments on a monthly level. This sample consists of 10,266 individuals from West Germany, since administrative data on East Germany (1,809 individuals) is only available starting in 1990 and therefore not used. The sample consists of about as many women as men for the birth years 1944 to 1989 (Table 1.3). The final samples used are described in the following section.

It is important to note that, although for some cohorts almost the entire working life is observed, how far earnings trajectories can be observed depends on the birth year of an individual. Since life-cycle employment patterns are not derived for future periods individuals can only be observed up until the age with existent data. Thus, it should be noted that for younger age cohorts "lifecycle" earnings are incomplete and should not be compared to those of older cohorts.

1.3.2 Empirical proceeding

Table 1.3 shows two different samples. Both are created by constructing yearly earnings for every individual and ages 18 to 64 that fall into the years 1975 to 2011. The first sample (EMPL) solely relies on information from employment spells. This means

that earnings are mostly positive.[9] While the first sample is restricted to spells where individuals are in an employment (EMPL), the second sample includes periods where individuals have no earnings in addition to this (EMPL+). The latter are panels where ages without social security records are included with earnings of zero for every person-age between 18 and 64. Neither of the samples is restricted to full-time or full-year employments. The relevance of EMPL+ trajectories is higher for groups where labor supply changes to a large extent over the course of life. Including periods for individuals who are not in the labor force is seldom done in the literature, one study where these are included in the calculation of work-life earnings is Tamborini et al. (2015).

Constructing age-earnings profiles is as follows: Using gross daily wages[10], imputed wages are used to calculate yearly earnings. This is done by summing up the days in employment in a year and multiplying these with the corresponding daily wages. Yearly earnings are deflated with a 2010 base year Consumer Price Index and weights are generated and applied in order to attain representative results.[11] The age-earnings profiles are constructed by calculating the unweighted arithmetic means and medians of annual earnings by age, separately according to educational levels. Work-life earnings are the sum of yearly earnings during working ages. The present value (PV) of work-life earnings is calculated by applying a 2 percent discount factor and calculating the sum of annual discounted earnings to the year a person turns 18. The same discount factor is also applied to annual earnings for the construction of age-earnings profiles.[12]

[9] Zero wages are only included if these are employment interruption notifications (e.g. maternity leave, longer periods of sickness).

[10] These also cover the social security insurance components pension, health, unemployment, nursing care and accident insurance.

[11] Weights are generated by using gender, region, education and birth cohort from the NEPS survey adjusted to the marginal distribution of the Mikrozensus 2010 by the German Federal Statistical Office.

[12] As an alternative, the inflation adjusted 10-year government bond yield is also used. While the 2 percent discount factor is assumed to be the same for all individuals irrespective of their age and the year they turn 18, the government bond yield implies differing discount factors for individuals (government bond yield the year a person turns 18). However, for both methods it is assumed that the discount factor is constant over the course of life of an individual. The consideration behind these assumptions is that using government bonds to calculate present values of work-life earnings is a more accurate model of the investment decision of an individual at age 18. Between 1975-1995 the nominal government bond yield is very high and on average 7.4 percent which results in very low work-life earnings and less informative patterns of age-earnings profiles than without discounting. The results are not displayed here but are available from the author upon request.

Education is classified into low, medium and high skilled (see Section 1.A.3) to best reflect the German education system, where apprenticeships traditionally play an important role in the vocational training system. Low skilled are defined as not having completed an apprenticeship or any other post-secondary education and as not having completed a high school degree. Medium skilled have an apprenticeship or a high school degree, while high skilled have a university or university of applied science degree. In the data, the actual average starting age of employment for men is 18.6 for low skilled (only individuals under the age of 25 considered), 19.6 for medium skilled (only individuals under the age of 30 considered) and 24.6 for high skilled (all individuals considered). For the age-earnings profiles the ages of starting an employment are therefore set to ages 18 for low skilled, 20 for medium skilled and 25 for high skilled and earnings are assumed to be zero before that.

1.4 Analysis

1.4.1 Patterns of age-earnings profiles

Figure 1.A.1 documents different age-earnings profiles for educational levels calculated by annual mean earnings for both samples. Sample EMPL for men in (a) and EMPL women in (b). Discounted with a 2 percent discount factor for EMPL men in (c) and EMPL women in (d). The EMPL+ sample include periods with zero earnings for men undiscounted in (e) and discounted in (g). For EMPL+ women these are in (f) and in (h).

1.4.1.1 Employed (EMPL) sample

In the EMPL sample the tradeoff of the human capital investment decision can be seen in Figure 1.A.1a for men. The age-earnings profile of a high skilled male worker is much steeper than the profile of the medium skilled worker and the high skilled worker sacrifices earnings early in the career to receive higher earnings in the future. High skilled men earn up to 72,408 EUR annually at the age of 48 and have an average annual earnings growth of 9 percent between the ages 25 and 40. While profiles for high skilled men follow an inverse u-shape this can not generally be said for profiles of medium and low skilled. The flat patterns for age-earnings profiles of medium and low skilled are evidence for lower growth rates throughout the course of working life.

As human capital theory predicts the yearly earnings increase with experience but at a diminishing rate. For men average annual earnings growth between ages 20-35 is 7.9 percent for medium skilled and 5.3 percent for low skilled. While the earnings growth is especially large for medium and low skilled men before the age of 30 and for high skilled men before the age of 40 it steadily diminishes thereafter. The increase in earnings of employed high skilled men (Figure 1.A.1a) for ages over 60 years could be explained by the drop in yearly average earnings for individuals in their mid-fifties. Individuals with very high earnings seem to continue working after the age of 60, while lower earning peers work less. This causes the remaining sample of the employed to be more selective.

Risk plays an important role when calculating work-life earnings due to uncertain career paths and possible shocks. Dillon (2018) reports that a worker is willing to give up at least 9 percent of her lifetime earnings to move from the most uncertain to the safest occupation at the start of her career. Different educational levels also lead to career paths with varying unemployment, wage and other risk. Higher education in most cases offers higher work-life earnings and therefore also a better protection against unemployment spells as can be seen in Figure 1.A.1a. However, risk-averse individuals who choose this career path also face foregone earnings for a longer period of time. These foregone earnings are calculated in subsequent sections.

The profiles of women in Figure 1.A.1b do not depict the inverse u-shape pattern seen for high skilled men since annual earnings in all skill groups stay on a similar level between ages 30 and 60. The highest annual earnings are reported for high skilled women at age 52 with 39,154 EUR. The medium skilled group has the highest number of observations and therefore offers the most reliable results. The group shows a low annual earnings growth. Earnings increase only at a decreasing rate already early in the career for reasons such as maternity. For this reason the more extensive second sample (EMPL+) used below is more informative for age-earnings profiles of women.

Discounted age-earnings profiles for men (Figure 1.A.1c) clearly show that earnings at older ages are less valuable and that starting to work at younger ages with lower levels of skill becomes more of an advantage with increasing interest rates. The highest present value of an annual earning is then at a lower age. For high skilled men the highest annual earnings are now at age 43 and are 43,582 EUR. For women the discounted profiles in Figure 1.A.1d are at low levels in all education groups with 24,639 EUR being the highest annual earnings for high skilled women at an early age of 31 years.

Comparing these results to other studies using German data, Bönke et al. (2015) find similar patterns for pooled birth cohorts 1935-49, as does the report by Schmillen and Stüber (2014) using earnings data from 2008-2010. However, these studies do not show profiles separately for men and women. They also rely on educational information from social security records known to be less accurate (Kruppe and Unger, 2014; Fitzenberger et al., 2005). Literature for the US, i.e. Tamborini et al. (2015), find similar differences between education profiles of men and women and patterns over the course of life.

Since the profiles constructed by annual mean earnings rely on the imputation strategy for top coded earnings, as a robustness check the profiles with annual median earnings are constructed. These are depicted for men in the EMPL sample in Figure 1.A.2a and suggest similar patterns of profiles.

1.4.1.2 Employed plus (EMPL+) sample

The EMPL+ profiles (Figure 1.A.1e, f, g and h), which refer to the sample that include periods with and without employment and zero earnings, are more compressed compared to the EMPL profiles. The average difference in annual earnings between the EMPL and the EMPL+ profiles are the lowest between age 45 and 50 in all skill groups for men. The highest annual earnings of high skilled men is 57,369 EUR at age 45 (Figure 1.A.1e). There is a steep drop in average annual earnings in the EMPL+ sample profiles for individuals in their mid-fifties indicating a lower participation in the labor force. Therefore, comparing the EMPL+ with the EMPL sample reveals that employment probabilities for men and women and actual labor supply differ to a certain extent. This is especially pronounced at the start and end of the career and could be due to various reasons. These include investing in education, early retirement as well as unemployment. EMPL and EMPL+ profiles for women both include periods of maternity leave which explains the similar pattern of profiles in Figure 1.A.1b, and f of medium skilled women. The profiles for women are much lower compared to those of men, however, relative differences between skill groups are quite similar to those of men.

Medium and high skilled profiles intersect in all graphs of Figure 1.A.1 and allow inference about the time needed to recover foregone earnings during education. Regardless of gender, discounting (0 percent, 2 percent) and sample restrictions (EMPL, EMPL+), high skilled individuals always manage to catch up in annual earnings with

low and medium skilled by the age of 30. High skilled and employed men are the most advantaged because they already have approximately twice the annual earnings in comparison to the low skilled at the age of 30.

As above with the EMPL sample, profiles with median annual earnings are constructed for men in the EMPL+ sample in Figure 1.A.2b. These show that in the high skilled group many individuals have very low or zero annual earnings in their twenties and towards the end of their fifties.

Bönke et al. (2015) and Schmillen and Stüber (2014) do not report age-earnings profiles of individuals that are not continuously employed and also do not report the profiles of women. Therefore our results complement both studies in adding zero earnings and suggest that average work-life earnings in the population are lower than previously suggested due to varying labor supply especially at the beginning and end of the career.

1.4.2 Cumulative earnings

The cumulative earnings give information on the age of a higher skill group intersecting with a lower skill group in cumulative work-life earnings. These intersection ages are interesting and help to verify the theory of human capital investment. For instance, they can answer several questions: How long does it take a high skilled person to make up for foregone earnings of education? What do these intersection points say about risk preferences?

Figure 1.A.2c and d are cumulative age-earnings profiles for men. While for the EMPL sample for men the graphs are almost linear for ages above 30, for the EMPL+ sample for men profiles seem nonlinear and show the very low earnings growth at the beginning and end of working life in all education groups.

Table 1.A.3 offers an overview of work-life earnings up to certain ages, i.e. 40, 50 and 60 years. These are also discounted with a 0 and 2 percent discount factor. The calculations show that for men and women earnings growth between 40 to 50 and 50 to 60 is similar in all skill groups. For men in both samples, higher skilled men always have higher cumulative earnings than lower skilled men. In the EMPL women sample this is also the case. However, it is not the case in the EMPL+ sample for women with a 2 percent discount factor. Here, high skilled women have earned less than medium skilled women at age 40 and age 50. In this case the human capital investment pays off very late, at age 56.

The intersection ages of cumulative earnings are depicted in Table 1.A.4. For high skilled employed men it takes until the age of 36 to catch up in cumulative earnings with the medium skilled. Taking into account periods without an employment increases this age to 39. For women, the EMPL+ sample provides a more realistic calculation of work-life earnings since the inclusion of zero earnings seems to be even more important than for men. This leads to an intersection age for women of 48 years and indicates that it takes a considerable amount of time to make up for foregone earnings. Using the 2 percent discount factor further increases these ages and reveals that the right human capital investment decision is not necessarily always higher education when periods without earnings are included and individuals retire early. In addition to this, direct costs of education, which are not included in this analysis, might make investment in education even less attractive. Late intersection points between some skill groups generally make investment in education riskier. The reason for this is that negative earnings shocks are more likely to impact work-life earnings before this intersection point. This is more likely to happen for women than for men since high compared to medium skilled women have relatively high intersection ages.

1.4.3 Work-Life earnings

The work-life earnings estimates are given in Table 1.A.5. These are calculated as a sum of annual average or median earnings for samples EMPL and EMPL+, men and women and educational groups. Median work-life earnings are calculated in addition to averages so as to have a robust estimation of work-life earnings that does not rely on the imputation strategy for top coded earnings. Results confirm the large impact of including zero earnings as was illustrated in age-earnings profiles. The work-life earnings calculated by means and medians and pooled for men and women do not differ immensely for the EMPL sample. However, the difference is far larger for EMPL+ since the median work-life earnings reveal the large number of individuals with low or no earnings at the start and end of working life. For the EMPL sample the earnings of high skilled men over the course of working life are 2.3 million EUR. For the EMPL+ sample, work-life earnings for high skilled men are 1.63 million EUR. This is 70 percent of EMPL work-life earnings and suggests that men with multiple periods of employment are outdistanced in terms of lifetime earnings from men who are employed throughout

their working life. Indeed, Johnson and Feng (2013) find that in the US 6 months after a job loss about half of long-term unemployed workers still have 40 percent lower incomes.

An important finding in Table 1.A.5 is that the sample restrictions make a considerable difference in work-life earnings which is especially pronounced for women. Work-life earnings of employed women (undiscounted) are 1.29 million EUR in the EMPL women sample and 0.7 million EUR for high skilled EMPL+ women sample. Considering the previously found strong relationship between labor force participation and earnings, the EMPL+ sample yields work-life earnings that are closer to the actual earnings in the population and should therefore be the preferred sample. In the EMPL+ sample for women, low skilled have work-life earnings of 0.43 million EUR and medium skilled have 0.62 million EUR. This means that, for low skilled women, work-life earnings in the EMPL sample are double that of the EMPL+. Discounting has a strong impact on the calculations with work-life earnings dropping from 1.63 to 0.96 million EUR for men and from 0.70 to 0.42 million EUR for women. This is similar in the other skill groups.

1.4.4 Foregone earnings of education

Measuring the implicit cost of education in form of foregone earnings requires a calculation of earnings of the next-best alternative. Here the next-best alternative for a person with a university degree (high skilled) is assumed to be the earnings of a person with an apprenticeship or high school qualification (medium skilled). In addition, one has to distinguish between two types of foregone earnings. First of all, there are foregone earnings that occur during education. Secondly, there are foregone earnings a person has after labor market entry. These are due to lower earnings which result from having less labor market experience than a person in a lower skill group.

The results are depicted in Table 1.A.6. Focusing on the EMPL+ sample, for high skilled men, foregone earnings during education are 64,880 EUR. These are the annual earnings medium skilled men receive up to the age of 25. This absolute value roughly corresponds to the annual earnings of high skilled men at age 40. To know how high these absolute values are as a fraction of work-life earnings is interesting because they can indicate whether human capital investment is profitable or not. This corresponds to about 4 percent of total work-life earnings for high skilled men. Taking into account the foregone earnings after labor market entry, i.e. up to the age at which high skilled earn

more than medium skilled for the first time, increases foregone earnings for this group to 109,556 EUR or 6.74 percent of work-life earnings. These are the sum of earnings of medium skilled up to age 25 plus the amount medium skilled earn more annually in the following 2 years.

For high skilled women in the EMPL+ sample foregone earnings are 65,638 EUR or 9.34 percent of work-life earnings. Including foregone earnings after starting to work increases these to 88,397 EUR or 12.58 percent. This is almost double the percentage of that in the EMPL sample indicating that foregone earnings are underestimated when not applying the correct sample restrictions. Foregone earnings of medium skilled are very low because in the data they only have two more years of education than low skilled. At the same time foregone earnings after starting to work are almost nonexistent, due to the fact that the medium skilled almost immediately have higher earnings than the low skilled.

1.4.5 At what interest rate is less education favored?

The interest rate is calculated for which present values of average work-life earnings of high and medium skilled men are equal. To find this rate the present value of work-life earnings of high skilled as a function of the interest rate r is equated with the present value of work-life earnings of medium skilled and the equation is solved for r. The less restricted sample of EMPL+ men is used here. Results are that for interest rates higher than 7.7 percent, the medium skill education (vocational training) is preferable to the high skill education (university). As a comparison, Hanushek et al. (2016) find a discount rate of 9.6 percent when comparing vocational and general education.[13] The importance of foregone earnings in the human capital investment decision also increases with higher interest rates. These results are interesting because the EMPL+ sample restrictions show relatively low capital market interest rates necessary to make less education preferable.

[13] Usinge the EMPL sample for men gives an interest rate of 10.1 percent and for women 6.68 percent. In the EMPL+ sample for women the rate is 2.95 percent. The rate for men from the EMPL sample is therefore similar to the 9.6 percent reported by Hanushek et al. (2016).

1.4.6 Earnings gain from an extra year of education

How large is the gain g in earnings from an extra year of education? This can be calculated by using the "elaborate" method originally proposed by Becker (1964). This method compares work-life earnings streams of high (LE_{high}) and medium (LE_{medium}) skilled and divides by the number of years more that high skilled have spent in school:

$$g = \frac{LE_{high} - LE_{medium}}{LE_{medium}} \times \frac{1}{5} \tag{1.1}$$

Foregone earnings and direct costs are included, however, direct costs are assumed to be a public subsidy and set to zero. It is further assumed that both skill groups have the same retirement age and that high skilled invest 5 years more in their education than medium skilled and have zero earnings while studying. Table 1.A.7 displays the results. For men the gain is 6.7 percent (EMPL+) and 8.0 percent (EMPL) while for women it is much lower at 2.6 percent (EMPL+) and 5.4 percent (EMPL). Naturally, with increasing interest rates g decreases. It is important to note that these results can not be interpreted as the causal effect of education. Nevertheless, the results found for men fall into the range of estimates of lower and upper bounds of 4.8 and 11.7 percent of the return to an additional year of education found by Ichino and Winter-Ebmer (1999).

1.4.7 Sample restrictions

Concluding, as has become evident, employment probabilities seem to differ from labor supply which makes it important to include individuals that have unsteady employment biographies as was done in the EMPL+ sample where zero earnings are included. Since education also has a powerful effect on the risk of unemployment using this sample seems especially important when interested in the lifetime perspective of education and earnings. The only other study that uses similar sample restrictions is Tamborini et al. (2015) for the US. Interestingly, there, results between EMPL and EMPL+ are much more similar than in our study for Germany where results differ to a large extent. This may be due to a vast number of differences, e.g. social security systems and labor markets, between the countries. In the German context sample restrictions similar to the EMPL+ sample should, however, also be considered in future research.

1.4.8 Cohort analysis

In the data, earnings trajectories can be observed for the age range from

$$\max\{18, 1975 - \text{year of birth}\} \quad \text{to} \quad \min\{64, 2011 - \text{year of birth}\}.^{14}$$

Concerns in the analysis are therefore that real earnings between cohorts might differ to such a large extent that it is inappropriate to represent profiles in one graph. Furthermore, age and period could generally confound and fewer administrative records being available for older cohorts could bias the results.

In order to have a sufficient number of cases, birth cohort groups are defined by five or more adjacent birth years and are pooled for men and women. Table 1.A.8 reveals that the cumulative earnings up to the age of 40 do not differ substantially for individuals born before and after 1960. For the EMPL sample and high skilled the difference is 1.19 percent, for medium skilled 0.77 percent and for low skilled 4.26 percent. For the EMPL+ sample average work-life earnings of medium and high skilled born after 1960 are higher than of those born before 1960, for the low skilled the opposite is the case.[15]

Figure 1.A.3 gives an overview of age-earnings trajectories for nine birth cohorts of the EMPL sample. Skill groups are not considered in this graph as to have a sufficient number of cases for each cohort. While these profiles do differ especially for the ages between 25 and 40 there are no systematic cohort patterns. The same can be said in the EMPL+ sample in Figure 1.A.4.

Splitting ages into eight groups also permits a comparison of birth cohorts for skill groups at the mean of every age group. These graphs in Figure 1.A.5 show similar patterns but differ in years with a very low number of observations. It is not possible to calculate complete work-life earnings for all birth cohorts, due to the simple fact that future earnings for the young birth cohorts are unknown. However, the analysis of birth cohorts provides evidence that for the analyzed years work-life earnings of birth cohorts are similar.

[14] For example, individuals born in 1944 can be observed for ages 31-64 while the age range is 18-41 for the birth year 1970 and 18-25 for the birth year 1986.

[15] Among others, one explanation for the latter could be *skill-biased technological change*, a demand shift triggered by new technologies that increased the demand for higher skilled workers in relation to lower skilled workers. In this scenario there are more low skilled that are unable to find a job as they are not sufficiently qualified which explains the lower earnings of low skilled after 1960 compared to before 1960. The patterns of age-earnings profiles for these cohorts support this theory. These age-earnings profiles are not depicted here but are available from the author upon request.

1.5 Conclusion

This study draws attention to the lifetime perspective of education and earnings. The calculation of work-life earnings is performed by using an innovative dataset, namely survey data linked to administrative social security records. Thereby this study avoids the data limitations of similar studies which use cross-sectional data and less reliable educational information when calculating work-life earnings.

The study contributes to the literature, firstly, by describing patterns of age-earnings profiles for educational groups and calculating work-life earnings and foregone earnings during education which are opportunity costs of education. Secondly, it uses far-ranging longitudinal earnings data. Thirdly, it introduces a less restrictive sample that includes periods of individuals without employment as zero earnings.

Work-life earnings for high skilled men are on average 2.3 million EUR and for women 1.29 million EUR if they are employed continuously throughout their career. Using the less restrictive sample which includes periods without employment as zero earnings, results in work-life earnings of 1.6 million EUR for men and for 0.7 million EUR for women. Results also suggest that foregone earnings are important implicit costs in the human capital investment decision of individuals. How large foregone earnings are as a fraction of lifetime earnings, however, depends on the kind of degree, gender and whether a person is employed throughout his or her career. These results have a high policy relevance since expected work-life earnings and forgone earnings strongly influence the human capital investment decision as well as the decision to use student loan programs. Regarding the sample restrictions, results have provided evidence that work-life earnings are much lower when including periods of individuals without employment. This has rarely been done before and seems important since the assumption that individuals are employed throughout their career is unrealistic. Future research should therefore generally include subgroups of the population with unsteady labor force participation in the calculation of work-life earnings.

1.A Appendix

1.A.1 Imputation of censored wages

Censored wages in the administrative data are imputed for employees subject to social security contributions for the years 1975-2011. Marginally employed, which have only been recorded since 1999, are dropped. Furthermore, spells are split to a monthly level. Upper ceiling limits are calculated by deriving daily limits from the pension insurance yearly limits given in the appendix of the Sozialgesetzbuch VI. The limits vary for years, region (East/West Germany) and type of pension insurance (workers/employees vs. miners). Our imputation strategy is similar to that of related literature (Gartner, 2005; Dustmann et al., 2009; Reichelt, 2015) and involves estimating a truncated regression and predicting the right censored wages and adding an error term. This strategy rests on the assumption that log wages follow a normal distribution. The regression is then repeated for each year. The variables in the prediction include education, sex, age, number of employees of the establishment, German nationality, indicator for workplace in East Germany, occupational status and occupation with all information taken from the social security records. Our main methods are:

- Method 1: Replacing all values larger than and equal to the censoring limit with imputed values (In total 6.67 percent of monthly spells are censored).
- Method 2: Replacing all values 0.985 times larger than the censoring limit with imputed values (in total 8.65 percent of all monthly spells are censored).

The following histograms for the year 2009 show, that the imputation works best for the 10,266 individuals from West Germany using imputation method 2. This is why method 2 is used as the main method for the calculation of age-earnings profiles and work-life earnings. Using method 2 gives a wage distribution that is censored each year between 1-5 percent for the years 1975-1984, 5-10 percent for the years 1985-1994 and 8-13 percent for the years 1995-2011. Comparing this to the IABS, a 2 percent random sample of social security records for the years 1975-2004, Dustmann et al. (2009) report that each year 9.4% and 14.2% of the male wage distribution there is censored. The 1,809 person sample for East Germany suffers from a small sample size and is not used for the analysis. Using imputation method 2, years and regions are pooled and histograms for different age and education groups are compared. N denotes

the number of Observations, i.e. number of spells on monthly level. Due to the data protection legislation of the Institute for Employment Research it is not permitted to export histogram bins with less than 20 observations. The Education groups are: No Abitur without completed vocational training (NOVOC); No Abitur with completed vocational training (VOC); Abitur without completed vocational training (ABI); Abitur with completed vocational training (ABIVOC); Completion of a university of applied sciences (UAS); College/university degree (UNI).

Figure 1.A.6 show the distribution of the logarithm of daily wages before and after the imputation for the year 2009. Figure 1.A.7 and Figure 1.A.8 show the distribution for age groups and pooled for all years, while Figure 1.A.9 and Figure 1.A.10 show the same for the education groups.

1.A.2 Structual break in 1984 for one-time payments

One-time payments are only included starting from 1984 in the administrative data since commencing in the year 1984 these payments became subject to social security tax. Ignoring this can increase earnings inequality (Steiner and Wagner, 1998). One possible method of correction is to assume that quantiles above the median are most affected by the structural break. Regressions of the wage ratios between the years before and after the break on the percentiles can then be estimated and a correction factor can be calculated (Dustmann et al., 2009; Fitzenberger, 1999). However, the NEPS is somewhat different to the datasets used in the literature mentioned above. The oldest individuals in the NEPS are 40 years old in 1984. These individuals are aged 31 in the year 1975 when social security records start. This age effect is one possible explanation why there are almost no one-time payments that are larger than zero observed in the years directly after the structural break (in 1984, 1985 and 1986) but more starting in the 1990s. This stands in contrast to other studies, e.g. Fitzenberger (1999), who use the 1 percent random sample from German Social Security Accounts (IAB-Beschäftigtenstichprobe) with a different age composition of individuals. In our dataset, of the above mentioned 2.4 million monthly spells only 2859 have one-time payments that are larger than zero. That is only 0.1 percent. Since there are almost no one-time payments observed directly after the structural break in 1984, for this analysis no correction factor is applied.

1.A.3 Definition of the education variable

Label	CASMIN classification	Description
Low skilled	$1a/1b/2b$	No apprenticeship
Medium skilled	$1c/2a/2c\_gen/2c\_voc$	Apprenticeship or high school degree
High skilled	$3a/3b$	University (of applied science) degree

1.A.4 Tables and Graphs

Table 1.A.1: Sample selection

	(1)	(2)	(3)	(4)
			Heckman selection model	
	OLS Full Sample	OLS Selected Sample	Maximum likelihood	Two-step model
Selection equation				
Std.dev. of interviewer months for waves			-0.2277***	-0.2277***
			(0.0301)	(0.0301)
Covariates from wage equation			Yes	Yes
ρ			0.0202	0.2145
			(0.0777)	
Wald test (p-value)			0.81	
Wage equation				
Female	-0.0558***	-0.0559***	-0.0556***	-0.0552***
	(0.0082)	(0.0083)	(0.0083)	(0.0084)
Married	0.4998***	0.4938***	0.4944***	0.4970***
	(0.0089)	(0.0091)	(0.0092)	(0.0100)
Years of education	0.0779***	0.0760***	0.0761***	0.0767***
	(0.0018)	(0.0018)	(0.0018)	(0.0020)
Age	-0.0039***	-0.0040***	-0.0040***	-0.0039***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Observations	16450	15460	16861	16861

Notes: Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Wage refers to log net household income from NEPS survey.

Table 1.A.2: Means and t-test of individuals agreeing and disagreeing to linkage

	No consent for record linkage			Given consent for record linkage				
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Difference	t-statistics
Birth year	1,919	1963.55	11.49	15,205	1963.75	11.27	-0.203	(-0.74)
Female (==1)	1,925	0.5	0.50	15,215	0.51	0.50	-0.011	(-0.88)
Married (==1)	1,925	0.63	0.48	15,215	0.66	0.47	-0.037 **	(-3.26)
Years of education	1,880	13.92	2.39	15,006	13.95	2.33	-0.030	(-0.53)
Currently employed (==1)	1,925	0.77	0.42	15,215	0.78	0.41	-0.018	(-1.77)
Occupational status								
Employee (==1)	1,925	0.43	0.50	15,215	0.46	0.50	-0.026 *	(-2.15)
Worker (==1)	1,925	0.12	0.33	15,215	0.14	0.34	-0.014	(-1.70)
Self-Employed (==1)	1,925	0.12	0.33	15,215	0.10	0.29	0.027 ***	(3.69)
Household size	1,910	2.60	1.20	15,205	2.66	1.36	-0.066 *	(-2.02)
Mother: years of education	1,660	11.75	2.39	13,881	11.80	2.34	-0.054	(-0.88)
Father: years of education	1,674	12.85	2.45	13,999	12.87	2.40	-0.025	(-0.40)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.A.3: Work-life earnings up to age 40, 50 and 60

	Panel A		Panel B		Panel C	
	EMPL+	EMPL	EMPL+	EMPL	EMPL+	EMPL
Men						
r = 0.00						
Low skilled	0.384	0.533	0.593	0.822	0.765	1.139
Medium skilled	0.521	0.655	0.888	1.096	1.176	1.519
High skilled	0.542	0.759	1.101	1.464	1.539	2.106
r = 0.02						
Low skilled	0.297	0.418	0.419	0.585	0.502	0.737
Medium skilled	0.398	0.503	0.611	0.759	0.751	0.962
High skilled	0.392	0.556	0.716	0.965	0.928	1.272
Women						
r = 0.00						
Low skilled	0.188	0.372	0.313	0.578	0.419	0.788
Medium skilled	0.280	0.440	0.454	0.690	0.603	0.935
High skilled	0.246	0.461	0.467	0.825	0.660	1.175
r = 0.02						
Low skilled	0.150	0.294	0.222	0.414	0.273	0.514
Medium skilled	0.223	0.346	0.323	0.490	0.395	0.608
High skilled	0.182	0.343	0.310	0.554	0.403	0.721

Notes: Work-life earnings in million EUR. Notes from Table 1.A.5 apply here also. All calculations based on means. Panel A: Lifetime earnings up to age 40; Panel B: up to age 50; Panel C: up to age 60.

Table 1.A.4: Ages of intersection of cumulative work-life earnings

	High / Low		High / Medium		Medium / Low	
	EMPL+	EMPL	EMPL+	EMPL	EMPL+	EMPL
r = 0.00						
Pooled (means)	33	33	38	36	23	26
Pooled (medians)	28	26	30	28	20	22
Men (means)	34	33	39	36	24	27
Women (means)	33	34	48	38	22	25
r = 0.02						
Pooled (means)	33	34	40	37	23	26
Pooled (medians)	28	26	30	28	20	22
Men (means)	35	34	41	37	24	27
Women (means)	34	35	56	41	22	25

Notes: Notes from Table 1.A.5 apply here. High, medium and low refer to skill level. All calculations based on lifetime earnings with discount factor r. Reported are ages at which cumulative earnings of a higher skill group overtakes a lower skill group.

Table 1.A.5: Work-life earnings

	Low skilled		Medium skilled		High skilled	
	EMPL+	EMPL	EMPL+	EMPL	EMPL+	EMPL
r = 0.00						
Pooled (means)	0.56	1.03	0.92	1.39	1.34	2.08
Pooled (medians)	0.35	1.02	0.83	1.34	1.13	1.98
Men (means)	0.79	1.26	1.22	1.67	1.63	2.34
Women (means)	0.43	0.87	0.62	1.02	0.70	1.29
r = 0.02						
Pooled (means)	0.37	0.65	0.59	0.86	0.79	1.22
Pooled (medians)	0.22	0.65	0.54	0.84	0.68	1.17
Men (means)	0.51	0.79	0.77	1.03	0.96	1.37
Women (means)	0.28	0.55	0.40	0.64	0.42	0.77

Notes: Work-life earnings in million EUR for ages 18-64. Discount factor r. EMPL: Employments. EMPL+: Employments plus periods with zero earnings. Yearly earnings are set to zero for ages 18 and 19 for medium skilled and to zero for ages 18-24 for high skilled to account for different labor market entry ages. If yearly earnings are unavailable due to a low number of observations and the data protection legislation of the German Institute for Employment Research, the missing earnings are substituted by the last known annual earnings: For pooled (means and medians) the low skilled EMPL earnings for ages 61-64 are not available and are therefore set to the earnings at age 60. For the EMPL+ the earnings of low skilled men aged 63-64 are unavailable and set to that at age 62. For the EMPL low skilled men this is the case for ages 59-64. For EMPL high skilled men earnings at age 64 are missing and replaced by earnings at age 63. For EMPL low skilled women the same method applies for ages 61-64 and for EMPL high skilled women for ages 63-64.

Table 1.A.6: Foregone earnings

	Up to labor market entry				Incl. earnings after labor market entry			
	High skilled		Medium skilled		High skilled		Medium skilled	
	EUR	Percent	EUR	Percent	EUR	Percent	EUR	Percent
Men								
EMPL	91,458	3.90%	19,611	1.17%	102,303	4.37%	21,568	1.29%
EMPL+	64,880	3.99%	10,909	0.90%	109,556	6.74%	10,909	0.90%
Women								
EMPL	85,720	6.64%	17,089	1.68%	93,157	7.21%	17,089	1.68%
EMPL+	65,638	9.34%	10,262	1.65%	88,397	12.58%	10,262	1.65%

Notes: Earnings of high skilled compared to those of medium skilled. Medium skilled compared to low skilled. Percentages are given as a fraction of work-life earnings. In addition, Notes from Table 1.A.5 also apply here.

Table 1.A.7: Earnings gain for an additional year of education

	EMPL+	EMPL
r = 0.00		
Pooled (means)	9.0%	9.9%
Pooled (medians)	7.5%	9.5%
Men (means)	6.7%	8.0%
Women (means)	2.6%	5.4%
r = 0.02		
Pooled (means)	7.0%	8.3%
Pooled (medians)	5.1%	7.7%
Men (means)	5.1%	6.7%
Women (means)	0.9%	4.0%

Notes: Results refer to g from Equation 1.1.

Table 1.A.8: Cumulative earnings up to age 40

	Low skilled		Medium skilled		High skilled	
	EMPL+	EMPL	EMPL+	EMPL	EMPL+	EMPL
Born before 1960	280,235	468,142	382,860	565,909	417,961	683,148
Born after 1960	265,373	448,999	409,036	561,606	455,343	675,105

Notes: Earnings in EUR.

Figure 1.A.1: Age-earnings profiles (from mean annual earnings)

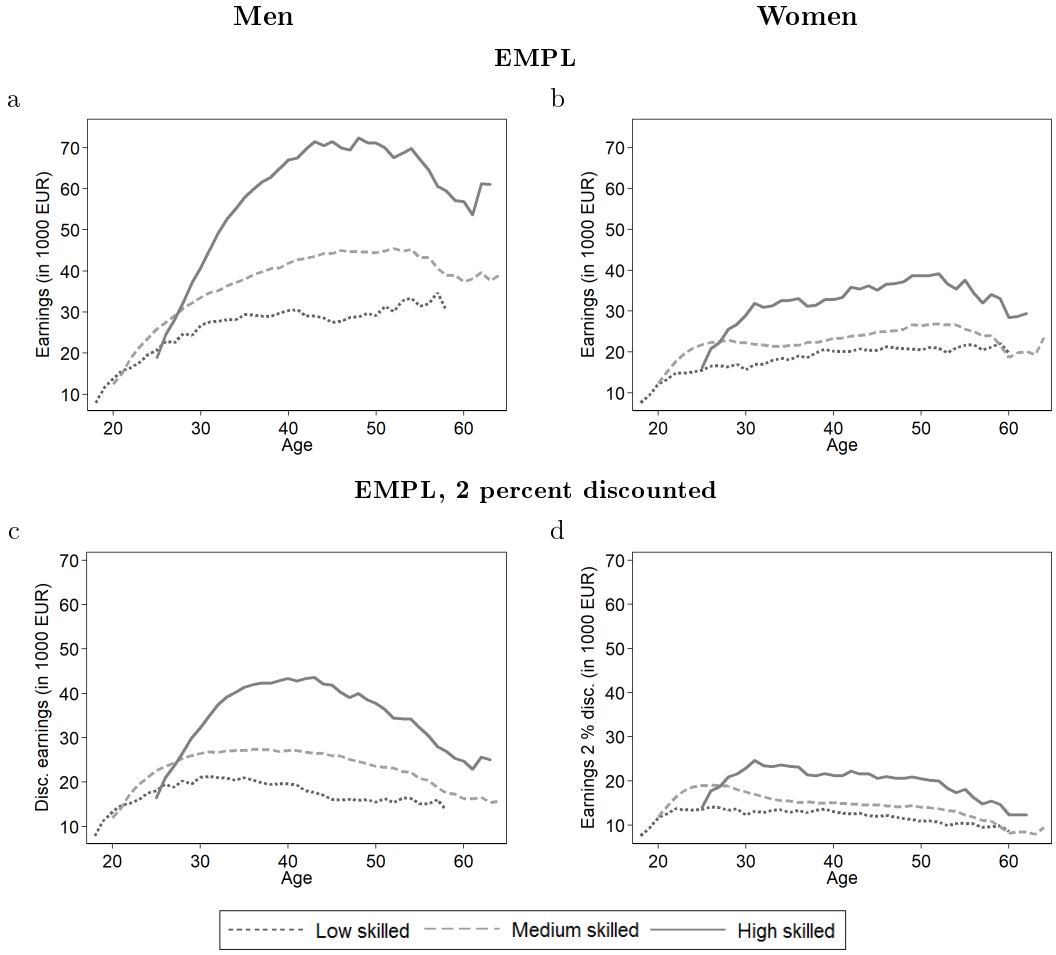


Figure 1.A.1 continued:

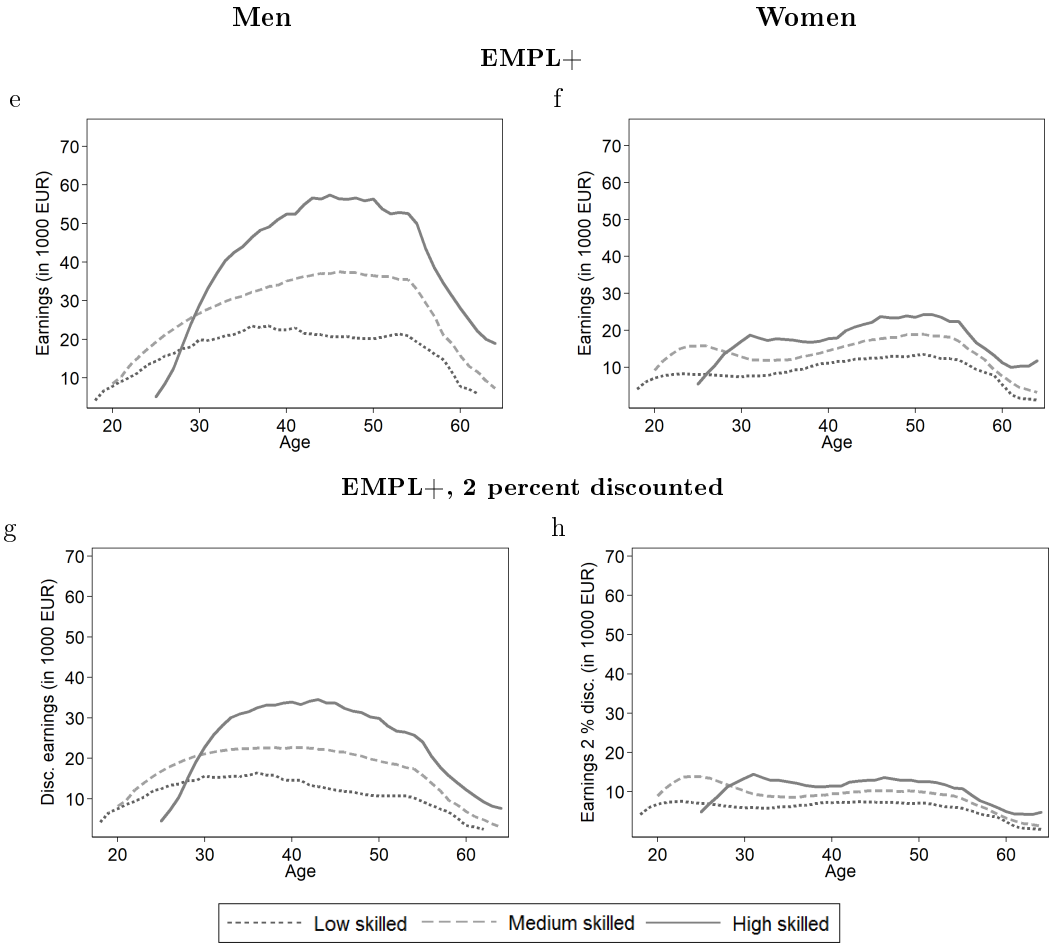
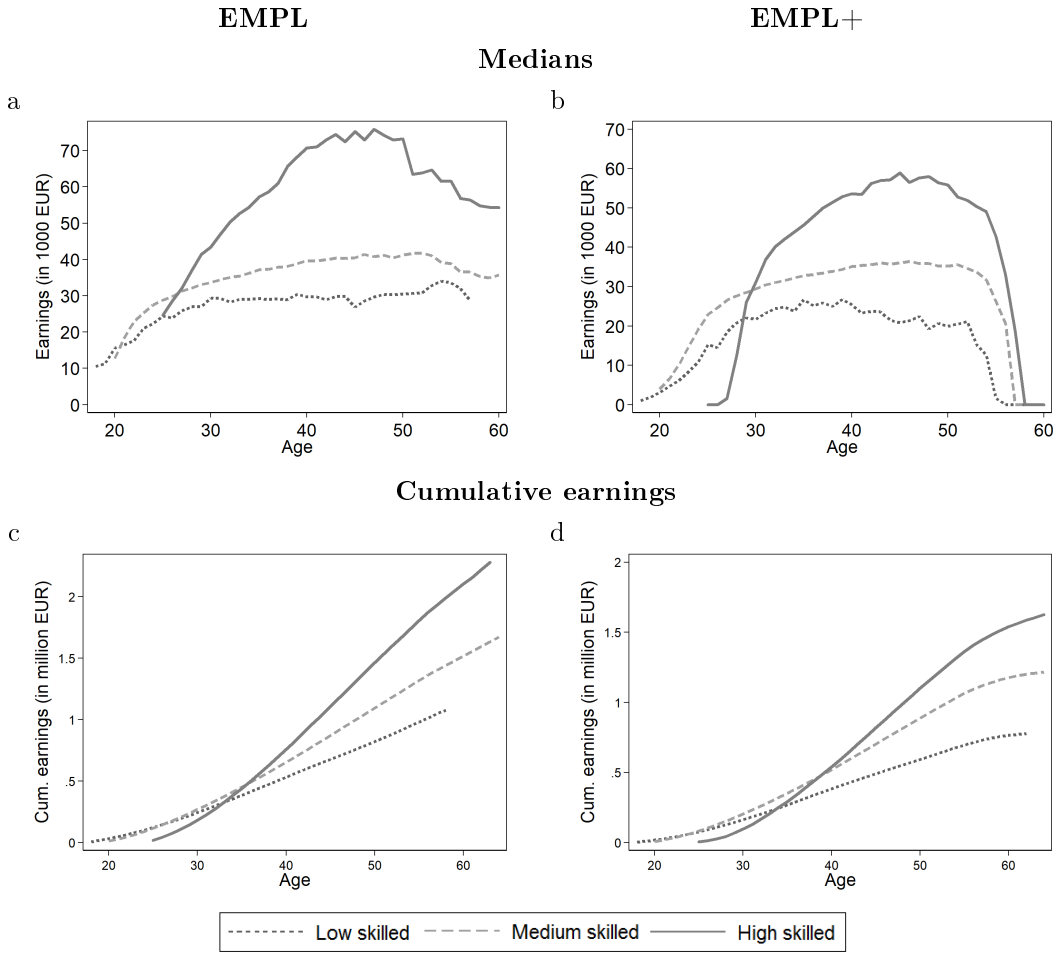
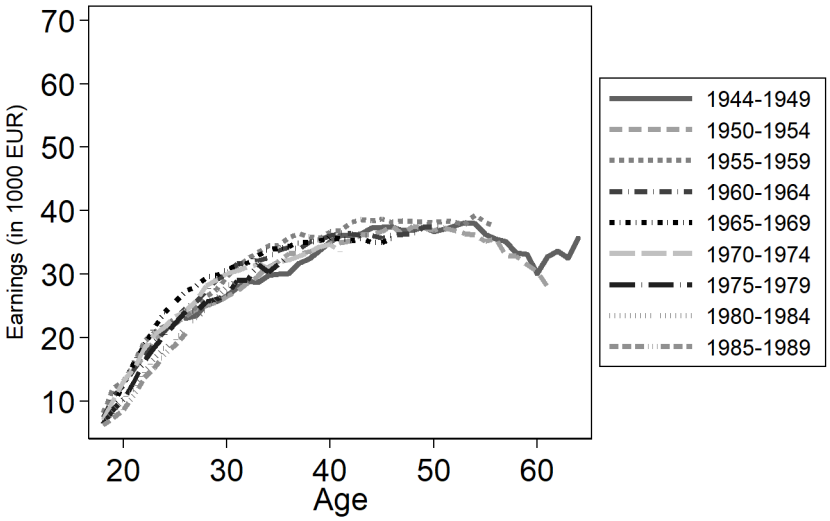


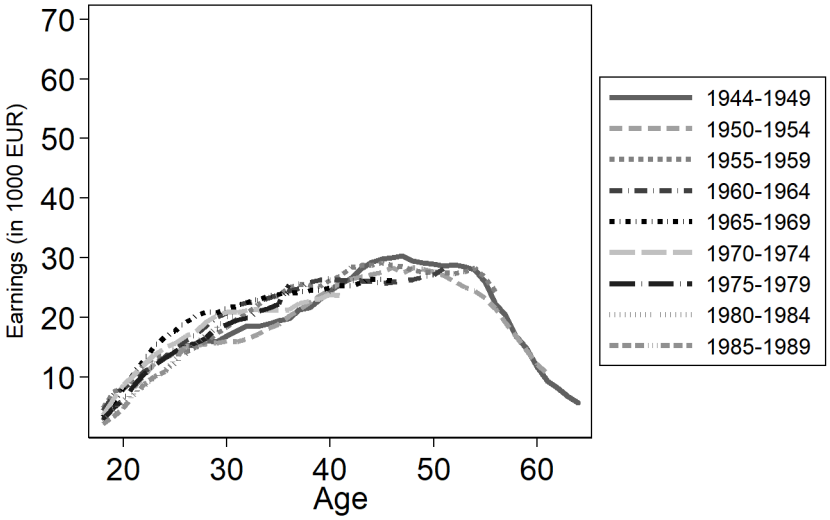
Figure 1.A.2: Age-earnings profiles for men





Notes: number of panel observations N: 1944-1949: N = 23110 ;
1950-1954: N = 27891 ; 1955-1959: N = 38032 ; 1960-1964: N = 40096 ;
1965-1969: N = 30973 ; 1970-1974: N = 15240 ; 1975-1979: N = 9308 ;
1980-1984: N = 6287 ; 1985-1989: N = 1992 ;

Figure 1.A.3: Age-earnings profiles for the EMPL sample for birth cohorts (pooled)



Notes: number of panel observations N: 1944-1949: N = 35520 ;
1950-1954: N = 42202 ; 1955-1959: N = 56446 ; 1960-1964: N = 59127 ;
1965-1969: N = 45499 ; 1970-1974: N = 23331 ; 1975-1979: N = 14767 ;
1980-1984: N = 10559 ; 1985-1989: N = 3348 ;

Figure 1.A.4: Age-earnings profiles for the EMPL+ sample for birth cohorts (pooled)

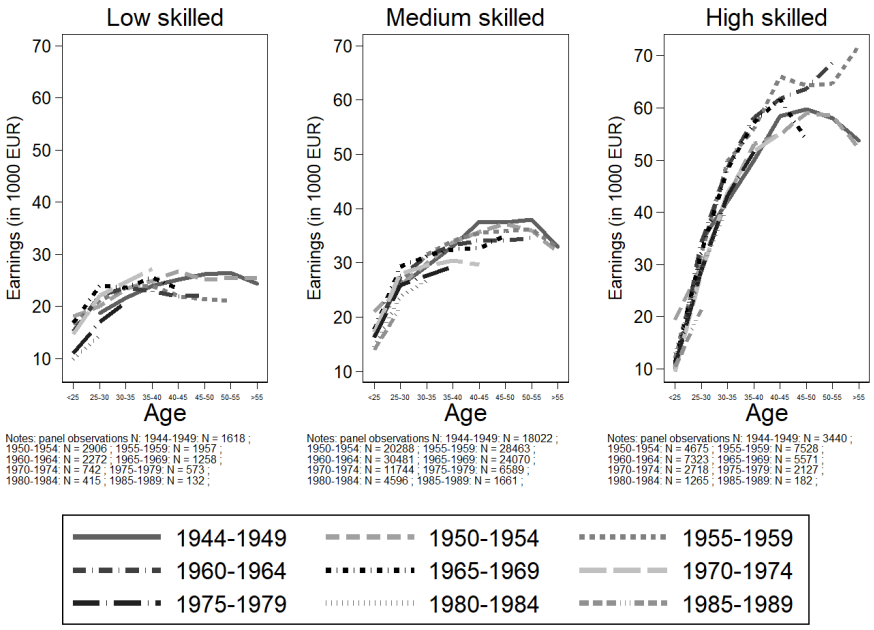


Figure 1.A.5: Age-earnings profiles for birth cohorts and skill groups (EMPL sample, pooled)

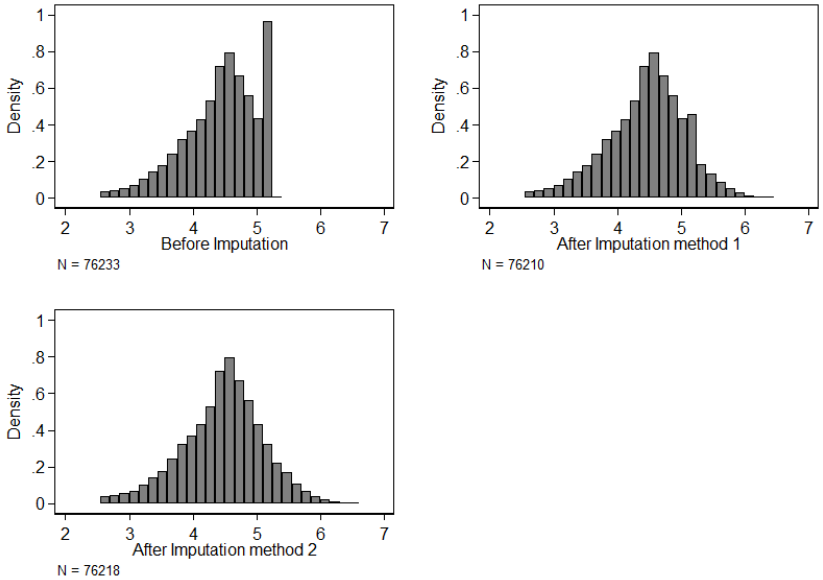


Figure 1.A.6: Distribution of logarithm of daily wage before and after the imputation (West Germany, year 2009)

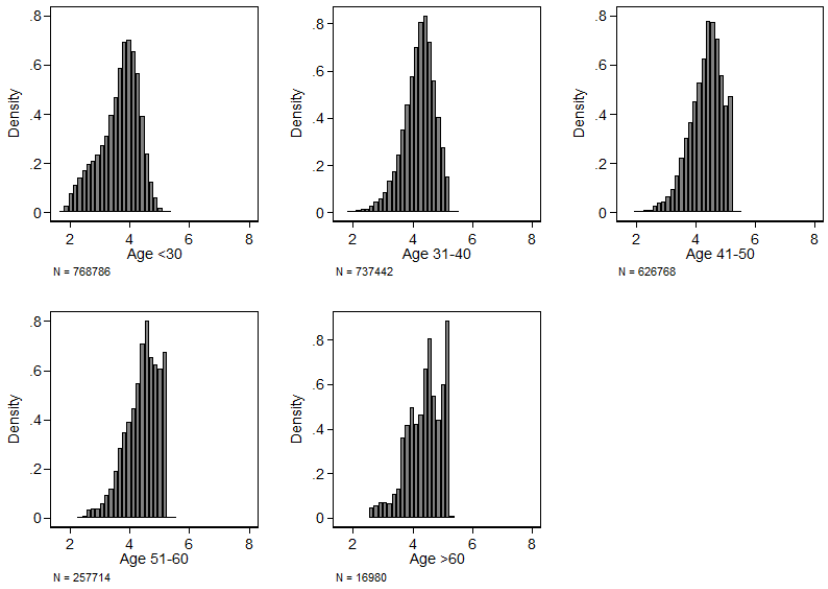


Figure 1.A.7: Distribution of logarithm of daily wage for age groups before the imputation

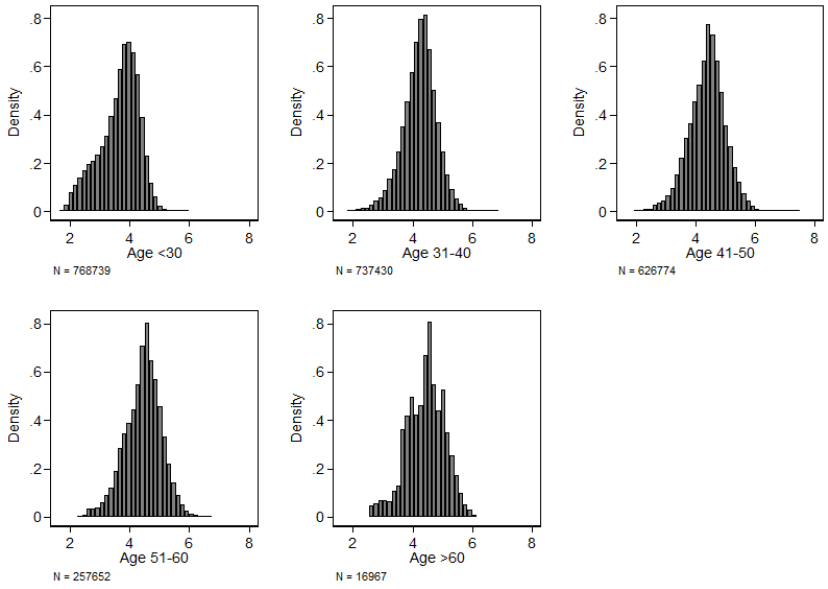


Figure 1.A.8: Distribution of logarithm of daily wage for age groups after the imputation

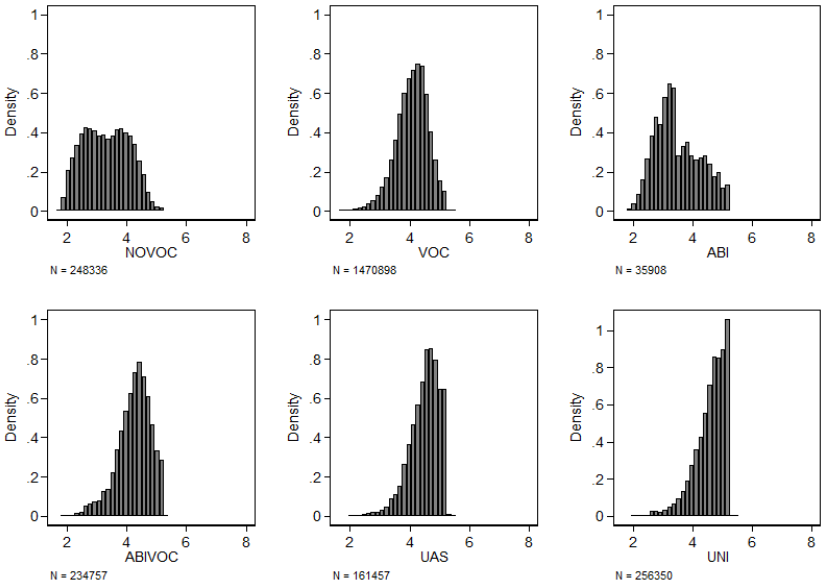


Figure 1.A.9: Distribution of logarithm of daily wage for education groups before the imputation

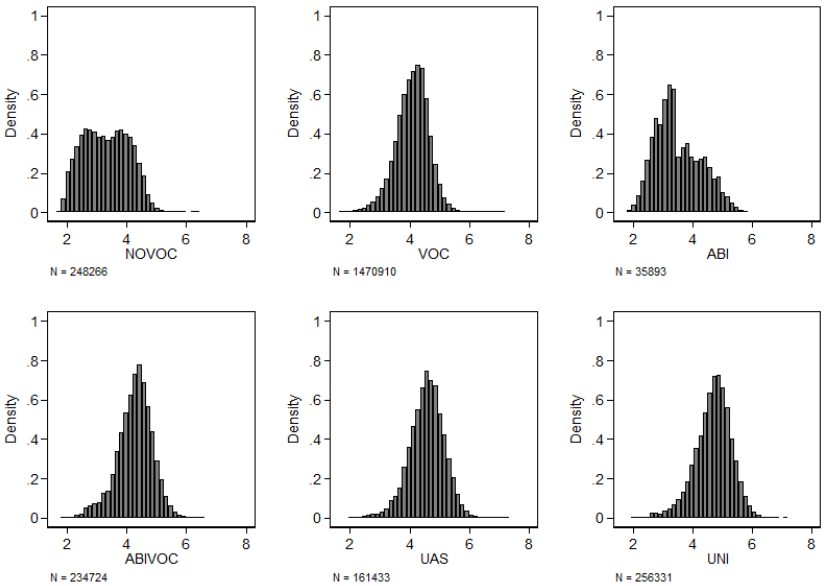


Figure 1.A.10: Distribution of logarithm of daily wage for education groups after the imputation

Chapter 2

Grades in Primary School and Educational Opportunity - An Empirical Analysis for Germany

2.1 Introduction

Should primary school reports include grades? While a large part of existing pedagogical literature is concerned with this question and focuses on soft skills such as the joy of studying, anxieties and self-concepts (Zeinz and Köller, 2006; Wagner and Valtin, 2003; van Aken et al., 1997), very little is known about the effects grading has on educational opportunity as measured by educational transitions and achievement for individuals with different parental background. This is interesting since there exists ample evidence for a strong relationship between parental background and secondary school track choice (PISA, 2007; Bos et al., 2007; Dustmann, 2004) as well as early tracking into differing-ability schools increasing this effect (Hanushek and Wößmann, 2006; Lange and von Werder, 2014; Brunello and Checchi, 2007)[1].

The aim of this empirical study is to contribute to the aforementioned literature by investigating, firstly, the effect of grading in the first classes of primary school on educational transitions and achievement and, secondly, on educational opportunity in an early tracking system. Hence, the results of this study are interesting for countries

[1] Even though recent evidence suggests that tracking does not have long-term effects since early track choices can be revised (Dustmann et al., 2017), these later career choices remain highly selective with respect to parental background (Biewen and Tapalaga, 2017).

that have early tracking systems (e.g. Germany and Austria) and/or countries where educational achievement is highly dependent on parental background (e.g. UK and US). For countries where grading does not start before class 6 (e.g. Sweden and Norway) and tracking is also done later, it is of interest to find out if later grading leads to a mismatch because school reports without grades are a less precise signal for students and parents. In Germany the question as to whether or not to include grades in school reports in primary schools has been heavily discussed since the 1970s when most federal states replaced grades in classes 1 and 2 by verbal assessments. Using these reforms this study will find out if children with low socio-economic status profit from later grading in terms of educational transitions and achievement. To do this, variation across West German federal states and across time is exploited and the fact that in the 1970s some federal states moved the time of grading in school reports to the third class while others still used grades before the third class. Using a difference-in-differences (DD) strategy the outcomes of these groups are compared. The data used in the analysis is survey data from the National Educational Panel Study (NEPS) for Germany and contains information about the educational biography of individuals and their socio-economic status. Findings show no effect of later grading on the likelihood of moving to upper secondary school for groups with different parental education. However, the reform increases later up- and downward transitions between secondary school tracks. This is evidence that verbal assessments might lead to a mismatching of individuals to secondary school tracks.

In theory there exist numerous ways in which later grading can, in interaction with school resources, peer effects, family resources and the belief a student has about his or her individual skills, have an effect on educational opportunity. One possible instance of this is that verbal reports are used by teachers to give distorted feedback to students with low self-confidence which could have a motivational effect. Bénabou and Tirole (2002) show in a theoretical model that one way of achieving a higher degree of self-confidence is to selectively process information about ability. Since there is a strong correlation between self-confidence and socio-economic background (Filippin and Paccagnella, 2012; Chowdry et al., 2011; Gregg and Washbrook, 2011; Chevalier et al., 2009) and initially small differences in self-confidence can lead to diverging patterns of human capital accumulation (Filippin and Paccagnella, 2012), using verbal reports to give distorted feedback could narrow this gap. Closely related to this is the

big-fish-little pond effect (Marsh and Parker, 1984)[2] which is increased by grading because social comparisons increase (Zeinz and Köller, 2006). The literature on rank finds that absolute grading motivates better than relative grading if students' abilities are independent (Dubey and Geanakoplos, 2010). At the same time high primary school (perceived) rank has positive effects on secondary school test scores (Murphy and Weinhardt, 2016). However, considering this, it is surprising that Wagner and Valtin (2003) do not find differences for the self-perception of abilities and performance motivation when comparing children with grades and verbal assessments but find that the joy of studying decreases for children with grades.

Teachers tend to have lower expectations for students with low socio-economic status (Figlio, 2005) and give lower grades to pupils with foreign origin (Sprietsma, 2013). This means that early grading could also lead to low achieving pupils being negatively labeled which in time could lead to further academic failure through self-fulfilling prophecies.[3] This raises the question as to whether an educational system with later grading can decrease the effects related to socio-economic status. As regards the family, it is likely that parents with different socio-economic status react differently to information from school reports. This is because cognitive costs, in general, are an essential part of decision making costs (Hastings and Tejeda-Ashton, 2008) and simplified information helps disadvantaged families to improve the educational achievement of their children (Hastings and Tejeda-Ashton, 2008). But Filippin and Paccagnella (2012) find that feedback has to be precise to help disadvantaged families. While grades are usually a precise signal for students and parents, a potential concern is that, were this not to be the case, mismatching to secondary school could occur. However, for any of the aforementioned ways educational opportunity is impacted it is required that, firstly, it makes a difference whether performance is judged by grades or verbal reports. Secondly, differences induced by not grading in classes 1 and 2 have to persist until the end of primary school.

The contributions of this paper are: First, providing causal evidence of the effect of having grades in school reports on educational transitions and achievement. This study is, to the best of my knowledge, the first to analyze the grading reform in the

[2]The effect that equally capable individuals have higher self-concepts when in a less capable group than in a more capable group.

[3]This is also known as the Pygmalion effect which was found by Rosenthal and Jacobson (1968) when conducting an experiment on self-fulfilling prophecy.

1970s for all West German states by combining retrospective educational biographies with rules on grading from law gazettes. Secondly, this study adds to the literature on ways to increase educational opportunity in educational systems where students are tracked into differing-ability schools at an early age and track choice is strongly related to socio-economic status.

The structure of this chapter is as follows: Section 2.2 presents an overview of the German education system, the primary school grading reform and the tracking system. The data is described in Section 2.3, the identification strategy is explained in Section 2.4 while Section 2.5 shows the results. Robustness checks of the results are provided in Section 2.6 and Section 2.7 concludes.

2.2 Institutional setting

2.2.1 Overview of the German school system

For the cohorts born between 1944 and 1987, which are the cohorts analyzed in the empirical section of this paper, compulsory schooling starts when children have reached the age of 6 and are enrolled into primary school (*Grundschule*). Primary school lasts 4 years in all West German states and at around the age of ten pupils are assigned to different types of secondary schools. The lowest secondary school track takes five years (*Hauptschule*), the middle track lasts 6 years (*Realschule*) and the highest track lasts for nine years (*Gymnasium*)[4]. Individuals who have attended the medium and low track usually proceed to a vocational education such as an apprenticeship in blue-collar or white-collar occupations. Individuals who have attained the final degree of upper secondary school (*Abitur*) usually proceed to university.

2.2.2 Primary school grading reform

For quite some time, but at least since the reformation five hundred years ago, grading is school has been a selectivity mechanism in meritocratic societes. In West Germany after 1950 school reports include grades already in the first classes of primary school.[5]

[4]These three school tracks are in the following referred to as lower (LS), middle (MS) and upper (US) school.

[5]However, most federal states evaluated the performance in the 1st class with one overall grade or grades in selected subjects and started grading all subjects in 2nd class.

State	BW			BY			HB			HH			HE			NI			NW			RP			SL			SH		
Class	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Year																														
1949/1950	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
...	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1970/1971	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1971/1972	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1972/1973	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1973/1974	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	V,G[b]	G[b]	G	G	G	G
1974/1975	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	V,G[b]	G[b]	G	G	G	G
1975/1976	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	V,G[b]	G[b]	G	G	G	G
1976/1977	G	G	G	G	G	G	V	V	G	G	G	G	G	G	G	G	G	G	V[a]	V[a]	G	G	G	G	V,G[b]	G[b]	G	G	G	G
1977/1978	V	V,G[e]	G	V[c]	G	G	V	V	G	G	G	G	G	G	G	V	V[c]	G	V[a]	V[a]	G	G	G	G	V,G[b]	G[b]	G	G	G	G
1978/1979	V	V,G[e]	G	V[c]	G	G	V	V	G	G	G	G	G	G	G	V	V[c]	G	V[a]	V[a]	G	V	V,G	G	V,G[b]	G[b]	G	V,G	G	G
1979/1980	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V,G	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1980/1981	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V,G	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1981/1982	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1982/1983	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1983/1984	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1984/1985	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1985/1986	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1986/1987	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1987/1988	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V,G	G	V,G	G	G	V,G	G	G
1988/1989	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V,G	G	G	V,G	G	G
1990/1991	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V,G	G	G	V,G	G	G
1991/1992	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V,G	G	G	V	V	G
1992/1993	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V,G	G	G	V	V	G
1993/1994	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V,G	G	G	V	V	G
1994/1995	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V	V	G	V	V	G
1995/1996	V	V,G[e]	G	V[c]	G	G	V	V,G[d]	G	V	V	G	V	V	G	V	V[c]	G	V	V	G	V	V	G	V	V	G	V	V	G

Notes: : after reform with grading starting before 3rd class : after reform with grading starting in 3rd class. V: verbal report. G: report with grades. V[a]: While the order is from the year 1979 it was already applied in 1976. V[b] Order from 1979 but already applied in 1973. V[c]: Two verbal reports. G[d]: grades in wording for reading, writing and Maths. G[e]: annual grading report only in maths and German. V,G: after first half of school year V and after second half of school year G. References: Law gazettes.

Table 2.1: Primary school grading reform

The 6-point grading scale ranges from 1 (excellent) to 6 (insufficient) and report cards at the end of a school year only allow for integer grades. In the 1970s a change in educational policy meant that recommendations were made for primary schools not to grade in classes 1 and 2 (KMK, 1970). Instead of grading in these classes a general verbal assessment of children was to be given. This should include an assessment of the social and work habits, indications of interests, special skills and weaknesses. In the following years most federal states started to follow these recommendations while others followed the recommendations only for the first class or up until the mid term report of the second class. Table 2.1 provides an overview of the year and extent of implementation of the primary school reform in states. While all states had semiannual grading reports at the latest in the third class, the states BW, BY, RP, SL and SH start grading before class three at least until the late 1980s. Table 2.1 also offers a yearly overview of the introduction of the reform in federal states with the differentiation stipulating whether grading started before or after the third class. The *light grey* area are states and years with grading before the third class after implementation of the reform. The *dark grey* area are states and years where grading was started in third class after reform implementation.

Verbal reports Reviewing the law gazettes of the federal states on the wording of the laws gives an idea of how comparable these actually were. In fact the wording in the law gazettes about the content of the verbal reports is very similar between states and follows the KMK ruling closely. Most of the time there are three main domains of assessment: Learning, working and behavior. Learning includes the learning progress, the capability of expression and speech comprehension. Working consists of working in class, in a group and alone. Behavior is assessed in respect to teachers, classmates and matters. Some states offer aid on formulating verbal reports by offering pre-built sentences where teachers only have to insert adjectives (Hartenstein and Ruddies, 1978). In addition to the orders of the federal states there exists a vast amount of literature on verbal reports offering aid as well as examples (Bartnitzky and Christiani, 1977; Hartenstein and Ruddies, 1978; Ullrich and Wöbcke, 1981; Heller and Bartnitzky, 1984). At the same time this literature also offers guidelines for teacher conferences and information on how to inform parents about the new reports. It is important to note that the substitution of reports with grades for verbal reports alone can not be seen as an improvement in performance assessment. If verbal reports are simply an

implementation of grades in words the traditional performance assessment will stay the same. Jachmann (2003) notes that for the reform to be successful there has to be a change in the culture of teaching and teachers as well as parents have to accept verbal assessments.

2.2.3 Tracking

Federal states have different rules when it comes to the assignment of children to secondary schools. While some states used to have entry exams as one criteria of transition to the higher secondary school track, the most important criteria is the grade point average as well as the teacher recommendation. The teacher recommendation is binding in most states in the 1950s, after which time this states gradually start to change this. This means that the influence of parents regarding the final decision of which secondary school to choose varies between states.

The educational transition to upper secondary school is still highly dependent on parental background (Dustmann, 2004). Tracking choices can, however, be revoked since it is possible to switch tracks at any given point during secondary school. Nonetheless, switching tracks before completion of a track seldom occurs (Dustmann et al., 2017). More frequently individuals change tracks after completing a track. The most common way for individuals in the low track to switch to the medium track is after grade 9 and for students of the medium track to switch to the high track after grade 10. Dustmann et al. (2017) find that by doing this earlier track choices are often revised, thereby avoiding any long-term consequences of track allocation. However, Biewen and Tapalaga (2017) show that most individuals that switch tracks have a high parental background.

In the following, an overview of the criteria for the transition from primary to secondary school is provided. Regarding grades and the teacher recommendation it is of particular interest to see whether these were binding or not. The rules are later used as control variables in the empirical analysis.

Teacher recommendation For the transition to secondary school the most important criteria is the recommendation of primary school teachers. While parent preferences and trial periods in higher secondary schools have also played a role in recent years, the teacher recommendation has for a long time dominated the transition process. Table 2.A.1 (Appendix) gives an overview of these rules. While Berlin was the only state

where the recommendation was not binding in the 1950s, most states started to enforce a system of non-binding recommendations in the 1970s and 80s. The southern states of BW and BY kept the binding recommendation as did NW. Jähnen and Helbig (2015) find evidence that the abolishment of binding recommendations led to an increase in the participation in higher education in Germany. This makes the rules all the more important to consider in an empirical analysis.

Grades Since 1950 some states have used grades given in the last class of primary school as a binding criteria to receive a recommendation for the highest secondary school track. This meant pupils had to receive certain grades or they would miss out on the chance of moving on to the highest secondary school type. Whether grades were binding or not can be seen in Table 2.A.2. While most states had non-binding grades up until the 1970s, some states such as BW, BY, HH, RP and SH chose to implement a system with binding grades from there on.

Formal tests From 1960 onwards, some states still used entrance examinations while others had already abolished these (KMK, 1960). The last state to have formal tests was Baden-Württemberg (BW) which had formal tests as the only criteria of transition to secondary school up until the school year 1966/67 and formal tests as one criteria amongst others up until 1978/79. These rules can be seen in Table 2.A.3.

2.3 Data and descriptive statistics

The analysis uses life-cycle data from the National Educational Panel Study (NEPS, starting cohort adults, SC6) as described in Blossfeld et al. (2011). The NEPS survey adult starting cohort was started in 2007 and covers 17,140 individuals born between 1944 and 1989. Up to the year 2015, seven NEPS waves are available which offer detailed information on educational biographies, family background and employment. The educational information consists of sequences of episodes of educational decisions. The retrospective collection of individuals' life courses in the NEPS can be seen as reliable since there is a data revision module.[6] In the analysis individuals are included that give

[6] After each interview a data revision is done and inconsistencies in the sequence of episodes are identified. The respondent is then asked to add missing events and dates (Skopek, 2013).

information about their primary school enrollment in West Germany (excluding West Berlin). Individuals are excluded that have a school starting age which is not between 4 and 9 years. Rudolf Steiner schools as well as other special schools are excluded, however, these represent relatively few cases. The main outcomes and characteristics

Table 2.2: Summary statistics of sample

Variable	Mean	Std. Dev.	Min.	Max.	N
Reform indicators					
D_1 (after reform: grading before 3rd class)	0.153	0.36	0	1	8444
D_2 (after reform: grading in 3rd class)	0.172	0.377	0	1	8444
Outcomes					
Transition from primary to upper sec. school (US)	0.417	0.493	0	1	8444
Years of schooling	11.338	1.742	7	13	8444
Track switch: Middle school (MS) $\rightarrow$ Upper sec. school (US)	0.064	0.245	0	1	8444
Track switch: Upper sec. school (US)$\rightarrow$ Middle/Lower school (MS/LS)	0.05	0.218	0	1	8444
Grade retention in 1-4 class	0.045	0.208	0	1	8444
Background variables					
Female	0.518	0.5	0	1	8444
Birth year	1966.31	10.216	1944	1987	8444
Age of primary school enrollment	6.504	0.511	4	9	8444
Age when starting sec. school	10.767	0.876	8	15	8444
Rules for transition to upper secondary school					
Binding recommendation	0.858	0.349	0	1	8192
Binding grades	0.294	0.456	0	1	8162
Entrance examinations exist	0.179	0.383	0	1	8168
Maximal education of parents					
Lower than vocational training (ED1)	0.082	0.275	0	1	8316
Vocational training, no upper secondary degree (ED2)	0.658	0.474	0	1	8316
Upper secondary degree (and possibly vocational training) (ED3)	0.079	0.27	0	1	8316
Tertiary education degree (ED4)	0.18	0.384	0	1	8316
Maximal occupational status of parents					
Low (OCC 1)	0.149	0.356	0	1	7666
Medium (OCC 2)	0.377	0.485	0	1	7666
High (OCC 3)	0.474	0.499	0	1	7666

Notes: Descriptives are unweighted. Occupational status: High: higher/lower managerial and professional workers, small self-employed with employees. Medium: routine clerical work, routine service and sales work, small self-employed without employees, manual supervisors, skilled manual workers. Low: semi- and unskilled manual workers, agricultural labor, self-employed farmers.

are listed in Table 2.2. The reform indicators are equal to one for states and periods that still started grading before 3rd class after the reform (D_1) or started grading in 3rd class after the reform (D_2). These indicators are constructed on the basis of

Table 2.3: Observations for federal states in the sample

Federal state	Freq.	Percent	Cum.
Schleswig-Holstein (SH)	312	3.69	3.69
Hamburg (HH)	183	2.17	5.86
Niedersachsen (NI)	1,136	13.45	19.32
Bremen (HB)	75	0.89	20.20
Nordrhein-Westfalen (NW)	2,540	30.08	50.28
Hessen (HE)	583	6.90	57.19
Rheinland-Pfalz (RP)	584	6.92	64.10
Baden-Württemberg (BW)	1,376	16.30	80.40
Bayern (BY)	1,488	17.62	98.02
Saarland (SL)	167	1.98	100.00
Total	8,444	100.00	

the school entry year of individuals and information on the grading reform from law gazettes. The key outcome variable is whether an individual transitioned to upper secondary school after having finished primary school. Other key dependent variables are the years of schooling[7], whether individuals switched school tracks and the probability of grade retention[8]. The final grade is only available for about half the sample with 1 being the best available grade and 5 the worst. The maximal education of parents is classified into four categories and occupations are classified into three categories. The parental education can also be grouped into two categories, i.e. with/without upper secondary degree (ED1/ED2 and ED3/ED4). Other variables are gender, birth year and the age of primary school and secondary school enrollment (decimal). While the fraction transitioning from primary to upper secondary school could be seen as rather high at 41.7 percent, one has to note that these statistics are unweighted.[9] The indi-

[7]The years of schooling are equivalent to the degrees which are 7 (*no degree*) , 9 (*Hauptschule*), 10 (*Realschule*) or 13 (*Abitur*). The total years of education is a function of the CASMIN classification. Using the educational attainment in years of schooling allows an easy to interpret continuous dependent variable to be constructed. However, when also looking at tertiary education it would be more appropriate to use a categorical variable instead of a continuous one since different educational paths can amount to equal years of education.

[8]This means repeating a grade.

[9]Applying weights reduces this value to 35.4 and is closer to the estimate found by Biewen and Tapalaga (2017) who use the same dataset but a much smaller sample. Weights are generated by using

cators for transition rules, as described in Section 2.2.3, are equal to one when at the time of transition to upper secondary school the teacher recommendation was binding, grades were binding or entrance examinations existed.

For the analysis observations from 10 West German federal states are used.

The identification strategy of the analysis rests on finding the correct state an individual went to school in. The NEPS offer variables on the federal state at birth as well as the state in which a person was enrolled into primary school. At the same time the dataset includes the residence history. To find the federal state at school start an imputation is done in the following order. First, the variable of the federal state of the primary school is used. Second, if this information is not available, the place of residence at the time of enrollment into primary school is substituted. Third, if the federal state is still missing the federal state at birth is applied. Table 2.3 shows that 77.45 percent of observations are from four states, these are NI, NW, BW and BY.

2.4 Identification strategy

Causal effects of the reform on educational outcomes are estimated by using a difference-in-differences framework. Two indicator variables are used to measure the effect of grading. Variable D_1 is equal to one for states where the primary school reform was introduced but where states still started grading before third class (e.g. in second class compared to grading in first class previously). Variable D_2 indicates grading in third class. Both indicators are equal to one for states and years effected by the reform.[10] The main specification is

$$Y_{ist} = \gamma_s + \lambda_t + \beta_1 D_{1,st} + \beta_2 D_{2,st} + X_{ist}\delta + R_{st}\mu + \phi_{st} + \varepsilon_{ist} \tag{2.1}$$

where Y_{ist} is one of the educational outcome variables for individual i in state s and time t. If the dependent variable is a binary variable such as an indicator for

gender, region, education and birth cohort from the NEPS survey adjusted to the marginal distribution of the Mikrozensus 2010 by the German Federal Statistical Office.

[10] Even though both indicators are included in the model, measuring the effect for D_2 is more interesting. The results are also easier to interpret since the entire control group was exposed to grading before third class while D_1 uses as a control group grading at the start of primary school before the reform as well as grading in third class after the reform. This way D_2 can measure the absolute effect of later grading.

the transition to upper secondary school, switching tracks or grade retention, a linear probability model (LPM) is used. γ_s and λ_t are time-invariant state and year effects, $D_{1,st}$ and $D_{2,st}$ are Dummy variables for the reform as defined above and X_{ist} are individual level covariates such as gender and the age when secondary school started. R_{st} are state and year specific rules for transition to upper secondary school (as discussed in Section 2.2.3), these are three Dummy variables which are one if grades are binding, the teacher recommendation is binding or entrance examinations exist and zero otherwise. ϕ_{st} adds state-specific trends before the reforms. ε_{ist} is an error term. Equation 2.2 includes indicators for the parental education of individuals. These are groups $ED2$, $ED3$ and $ED4$ and are equal to one if the maximal education of an individual's parent falls into one of the categories. $ED4$ is the highest parental education. $ED1$ serves as a reference category for the parental education. The reform variables are then interacted with parental education groups:

$$\begin{aligned} Y_{ist} = {} & \gamma_s + \lambda_t + \beta_1 D_{1,st} + \beta_2 D_{2,st} + \alpha_1 ED2_{ist} + \alpha_2 ED3_{ist} + \alpha_3 ED4_{ist} \\ & + \delta_1 D_{1,st} * ED2_{ist} + \delta_2 D_{1,st} * ED3_{ist} + \delta_3 D_{1,st} * ED4_{ist} \\ & + \delta_4 D_{2,st} * ED2_{ist} + \delta_5 D_{2,st} * ED3_{ist} + \delta_6 D_{2,st} * ED4_{ist} \\ & + X_{ist}\delta + R_{st}\mu + \phi_{st} + \varepsilon_{ist} \end{aligned} \tag{2.2}$$

The hypotheses are:

- H_1: The effect of later grading on educational outcomes is larger when grading starts in 3rd class compared to when grading starts before 3rd class after the reform ($\beta_2 > \beta_1$ in Equation 2.1).
- H_2: Later grading (in 3rd class) leads to more equal opportunity in educational career paths ($\beta_2 + \delta_4 > \beta_2 + \delta_5 > \beta_2 + \delta_6$ in Equation 2.2).

The common trends assumption is a key identifying assumption which states that in the absence of the reform, changes in outcomes between birth cohorts living in states that adopted the reform would have been the same as in states where there was no reform. While pre-treatment can be used to show that trends are the same before and after the reform, it is still difficult to verify. This is because, even with the same pre-trends, other policies can change at the same time. For example, states could also have changed teaching methods before the reform was introduced. However, the literature

suggests that teaching methods did not differ significantly in classes with grading and classes with verbal reports (Wagner and Valtin, 2003). The common trends assumption is further addressed in Section 2.6.

2.5 Empirical results

Results are from estimating Equation 2.1.[11] Table 2.4 reports the effects of transitioning from primary to upper secondary school on the primary school reform. The estimates include state and year effects, state-specific time trends for the time before the reform[12], other controls such as the school entry age and gender as well as the rules for transition to upper secondary school. These are indicators that equal one if school recommendations are binding, grades are binding and an entrance examination exists. Column (1) refers to Equation 2.1 while column (2) refers to the model from Equation 2.2. The coefficients are larger in size for D_2 than for D_1 which confirms H_1, i.e. it is more likely that effects of verbal reports can be found for the individuals in the classes that started grading in 3rd class. The results in column (1) are that starting grading in 3rd class (D_2) increases the likelihood of attending upper secondary school by 4 percentage points. For grading before 3rd class after the reform (D_1) it is about -1 percentage point. Although the coefficients are not significant which is likely to be due to the large standard errors of estimated coefficients in the medium-sized sample these estimates provide some evidence for an effect of later grading on the transition to upper secondary school.[13]

Unsurprisingly, the parental education variables added in (2) have a strong influence on secondary school track choice. This has been found in many previous studies related to educational opportunity and tracking (e.g., Dustmann, 2004). The interaction terms in (2) reveal how the later grading effect changes with increasing parental education level. Results for the more interesting case of D_2 are that the "grading gap" does not seem to change with parental education (H_2). For individuals with parental education ED2 it is 0.078 ($= 0.0609 + 0.0171$), for ED3 it is -0.009 ($= 0.0609 - 0.0699$) and for

[11] Since the standard errors often understate the standard deviation of the estimators (Bertrand et al., 2004) they are clustered at the state × year of birth level.

[12] This allows treatment and control states to follow different trends before the reform and can therefore also be seen as a check of the identification strategy.

[13] While this model is a Linear Probability Model (LPM), marginal effects of a probit model confirm these results with average marginal effects of -0.0016 for D_1 and 0.0537 for D_2.

Table 2.4: DD estimates, visiting upper secondary track on grading reform

	(1)	(2)
Grading reform:		
D_1 (after reform: grading before 3rd class)	-0.0107	-0.0063
	(0.0813)	(0.0875)
D_2 (after reform: grading in 3rd class)	0.0407	0.0609
	(0.0873)	(0.0956)
Parental Education (Reference: $ED1$ (Lower than vocational training)):		
$ED2$ (Vocational training, no upper secondary degree)		0.1277***
		(0.0219)
$ED3$ (Upper secondary degree (and possibly vocational training))		0.4295***
		(0.0311)
$ED4$ (Tertiary education degree)		0.5910***
		(0.0236)
Interaction of D_1 with $ED1 - ED4$ (Reference: $D_1 * ED1$):		
$D_1 \times ED2$		0.0522
		(0.0443)
$D_1 \times ED3$		-0.1191*
		(0.0681)
$D_1 \times ED4$		0.0653
		(0.0447)
Interaction of D_2 with $ED1 - ED4$ (Reference: $D_2 * ED1$):		
$D_2 \times ED2$		0.0171
		(0.0534)
$D_2 \times ED3$		-0.0699
		(0.0794)
$D_2 \times ED4$		0.0183
		(0.0531)
State effects	✓	✓
Year effects	✓	✓
State-specific trends	✓	✓
School entry age	✓	✓
Female	✓	✓
State and year specific rules for transition to upper secondary school:		
Binding recommendation	✓	✓
Binding grades	✓	✓
Entrance examinations exist	✓	✓
Observations	8169	8045

Notes: OLS regression. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, the robust standard errors (in parentheses) are adjusted for clusters at the state × year of birth level. Transition rules are implemented as Dummies (see Table 2.2).

ED4 it is 0.0792 ($= 0.0609 + 0.0183$). However, none of the interaction terms for D_2 are significant. The F-test of joint significance with the null hypothesis H_0: $\beta_2, \delta_4, \delta_5, \delta_6 = 0$ reports an F-statistic of 1.00 with a p-value of 0.406 and can not be rejected.[14] Separate tests with any of the interactions of D_2 and a given educational category all report an F-statistic of below 1.00 and therefore also cannot be rejected.

Nevertheless, the interaction terms do also indicate that there is some evidence of later grading having an effect on educational opportunity.

Table 2.5 give estimates for different educational outcomes. None of the coefficients for the years of schooling (2) are significant at the 10-percent level. While the coefficient for D_1 has a negative sign, the coefficient for D_2 is positive but very small (-0.0011). The likelihood of individuals switching tracks up or down at a later point are almost identical for D_2 and significant at the 10-percent level (columns (2) and (3)). There is no significant effect of later grading on the grade retention in class 1-4. However the coefficient for D_2 shows a negative sign indicating that later grading leads to a lower grade retention of 2.9 percentage points (4). This is mainly due to the fact that the grading reform also meant that generally there was no grade retention in classes that had reports that solely included verbal assessments. For the parental education of individuals, again, it becomes evident that both the years of schooling (1) and mobility between tracks (3) increase with parental education.[15] At the same time the likelihood of grade retention decreases (4). Regarding the interaction terms of the more interesting case of D_2, the coefficients indicate that the group with ED1 are likely to have profited the least from later grading. ED2 and ED3 are groups that had profited the most, however, the effects are not significant. None of the coefficients of the other outcomes are very large or significant in the case of D_2 interacted with parental education. Separately testing interactions of D_2 with education categories results in most tests not being rejected. However, for the downward school switch (3) all tests show F-statistics above 3 and p-values below 0.05. This suggests that while later grading might have caused some mismatch, this may also be connected to selectivity regarding parental background. The mismatch could be caused by reports without grades being a less precise signal, making sorting more difficult. This results in track choices being revised at a later

[14] The F-test of joint significance for the less interesting null hypothesis H_0: $\beta_1, \delta_1, \delta_2, \delta_3 = 0$ reports an F-statistic of 2.17 and a p-value of 0.0720.

[15] The latter has been thoroughly investigated by Biewen and Tapalaga (2017).

Table 2.5: DD estimates, educational outcomes on grading reform

	Years of schooling (1)	MS → US (2)	US → MS/LS (3)	Grade ret. 1-4 class (4)
Grading reform:				
D_1	-0.3748	0.0394	0.1091**	0.0498
	(0.4275)	(0.0514)	(0.0464)	(0.0574)
D_2	-0.0011	0.1168*	0.1212**	-0.0293
	(0.4393)	(0.0607)	(0.0489)	(0.0566)
Parental Education (Reference: ED1):				
$ED2$	0.7485***	0.0110	0.0155*	-0.0643***
	(0.0845)	(0.0110)	(0.0084)	(0.0145)
$ED3$	1.6658***	0.0110	0.0333**	-0.0589***
	(0.1162)	(0.0170)	(0.0145)	(0.0184)
$ED4$	2.1268***	-0.0007	0.0435***	-0.0702***
	(0.0863)	(0.0141)	(0.0133)	(0.0136)
Interaction of D_1 with $ED1 - ED4$ (Reference: $D_1 * ED1$):				
$D_1 \times ED2$	0.0954	0.0440*	0.0341*	-0.0449
	(0.1854)	(0.0224)	(0.0199)	(0.0383)
$D_1 \times ED3$	-0.2264	0.0989**	0.0647*	-0.0781*
	(0.2592)	(0.0418)	(0.0370)	(0.0405)
$D_1 \times ED4$	0.1951	0.0350	0.0242	-0.0616*
	(0.1926)	(0.0238)	(0.0269)	(0.0352)
Interaction of D_2 with $ED1 - ED4$ (Reference: $D_2 * ED1$):				
$D_2 \times ED2$	0.1314	-0.0102	0.0017	0.0018
	(0.1936)	(0.0269)	(0.0205)	(0.0420)
$D_2 \times ED3$	0.0611	-0.0219	-0.0076	-0.0383
	(0.2317)	(0.0402)	(0.0310)	(0.0422)
$D_2 \times ED4$	0.0408	-0.0180	-0.0266	-0.0514
	(0.2047)	(0.0317)	(0.0230)	(0.0394)
Observations	8045	8045	8045	8045

Notes: OLS regression. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, the robust standard errors (in parentheses) are adjusted for clusters at the state × year of birth level. All models include state and year effects and state specific trends, a dummy for female, school starting age and state and year specific rules for the transition to upper secondary school (recommendation, binding grades, entrance examinations).

point. Many of these individuals that are in the high school track decide to switch tracks because they do not want to go to university but start an apprenticeship training instead. Indeed, Biewen and Tapalaga (2017) show that, after completion of upper secondary school, the share of individuals starting an apprenticeship training is as large as the share of students that start to study.

Even though the binary variables of rules for the transition to upper secondary school (binding recommendations, binding grades, entrance examinations) were included as controls in the regressions, it is worth further investigating these. Results could indicate that verbal assessments in combination with certain transition rules have a significant effect on the educational career path in groups with different parental education. Interaction terms of the reform with these rules are included in Table 2.6. Column (1) is the original sample, column (2) is a sample that only includes parents without upper secondary degree. The interaction of the reform with entrance examination rules is omitted because of collinearity. The main dependent variable remains the likelihood to proceed to upper secondary school. None of the interaction coefficients show significance at the 10 percent level. Also, none of the F-tests with interactions of D_2 with a given transition rule are significant at the 10-percent level. There is also no clear pattern that would suggest that the reform in interaction with transition rules benefited individuals with low parental education.

2.6 Robustness

The identification assumption Since there are multiple treatment and control states and periods it is difficult to provide visual evidence of identical counterfactual trends in treatment and control states. A formal test is to interact the treatment variable with time dummies (Angrist and Pischke, 2008; Autor, 2003). [16] This results in q "leads" (anticipatory effects: $\alpha_{+1}, \alpha_{+2}, ..., \alpha_{+q}$) and m "lags" (post treatment effects: $\alpha_{-1}, \alpha_{-2}, ..., \alpha_{-m}$). α_0 refers to the year of adoption. While α corresponds to D_1 (grading before 3rd class), β does the same analog for D_2 (grading in 3rd class). The

[16] Alternatively, Mora and Reggio (2015) offer a new command framework which allows the evaluation of treatment effects under alternative assumptions.

Table 2.6: DD estimates, transition to upper secondary school on Reform interacted with binding secondary school recommendation and binding grades

	All	Parents with ED1, ED2 (without upper secondary degree)
	(1)	(2)
Grading reform:		
D_1	-0.0312	0.2048
	(0.2080)	(0.1983)
D_2	0.0579	0.1656
	(0.2048)	(0.1927)
Rules for transition to upper secondary school:		
$D_{Binding\ recommendation}$	0.0298	-0.0362
	(0.0745)	(0.0699)
$D_{Binding\ grades}$	0.0248	0.0966
	(0.1229)	(0.1003)
$D_{Entrance\ examinations\ exist}$	0.2471**	0.1790**
	(0.1017)	(0.0855)
Interactions:		
$D_1 \times D_{Binding\ recommendation}$	-0.0366	-0.0068
	(0.1175)	(0.1073)
$D_1 \times D_{Binding\ grades}$	0.0705	-0.2165
	(0.1663)	(0.1434)
$D_2 \times D_{Binding\ recommendation}$	-0.1444	0.0632
	(0.1079)	(0.1457)
$D_2 \times D_{Binding\ grades}$	0.0473	-0.1441
	(0.1802)	(0.1679)
Observations	8169	5975

Notes: OLS regression. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, the robust standard errors (in parentheses) are adjusted for clusters at the state × year of birth level. All models include state and year effects and state specific trends, a dummy for female, school starting age. Interaction of entrance examinations with Reform Dummies omitted.

equation is:

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=0}^{m} \alpha_{-\tau} D_{1,s,t-\tau} + \sum_{\tau=1}^{q} \alpha_{+\tau} D_{1,s,t+\tau} + \sum_{\tau=0}^{m} \beta_{-\tau} D_{2,s,t-\tau} + \sum_{\tau=1}^{q} \beta_{+\tau} D_{2,s,t+\tau} + X_{ist}\delta + R_{st}\mu + \phi_{st} + \varepsilon_{ist} \quad (2.3)$$

The outcome variable Y_{ist} is the likelihood to transition to upper secondary school, other variables are as defined in Equation 2.1. The DD assumption is tested as to whether coefficients on all leads of the treatment are zero. At the same time the pattern of the lagged effects is of interest in order to find out if the causal effect gets larger or smaller. D_1 and D_2 are estimated in one model as shown in Equation 2.3. The estimated leads and lags coefficients are plotted in Figure 2.A.1 and Figure 2.A.2 and run from five years ahead to five years behind. In Figure 2.A.1 for coefficients of D_1 the estimates before the reform show a clear pattern only starting in the year before the adoption. In Figure 2.A.2 for the coefficients of D_2 the pattern is consistent with a causal interpretation. That is, there are no large anticipatory effects on the likelihood of transitioning to upper secondary school. In the third year after the reform these effects strongly increase. To investigate the robustness of the results and whether the DD estimates are actually the effect of the later grading reform and not just state or year trends, a similar placebo test is performed with more years. A placebo reform indicator is constructed where the reform is shifted to earlier years. Table 2.7 gives the results of these estimations. The likelihood of attending upper secondary school for the entire sample on the D_2 placebo is small and insignificant. This shows that Placebo effects are weak. However, as before, the placebo does not work for every variable and year. The coefficients for D_2 are small and insignificant for the leads $t-7$ to $t-10$. Some of the coefficients are, however, larger and significant. Since year effects are included, the explanation that older cohorts have less education than younger cohorts does not hold.

In Columns (3)-(4) many of the coefficients are smaller than in Table 2.5 and mostly insignificant. The relatively large effect found previously of the reform on mobility between tracks cannot be found here. This speaks in favor of the robustness of the estimation. The coefficients for grade retention, on the other hand, are often higher in the placebo estimation for the D_1 placebo. This was due to the fact that it became less likely for children to repeat the first or second class as a result of laws being changed

Table 2.7: DD estimates, placebo test

	Transition to upp. sec. (1)	Years of schooling (2)	MS → US (3)	US → MS/LS (4)	Grade ret. 1-4 class (5)
D_1 Placebo (t-6)	0.0735	0.3162	-0.0522	-0.0057	-0.0727**
	(0.0735)	(0.3080)	(0.0332)	(0.0307)	(0.0349)
D_1 Placebo (t-7)	-0.0560	0.1303	0.0099	-0.0734**	-0.0272
	(0.0767)	(0.2282)	(0.0306)	(0.0347)	(0.0308)
D_1 Placebo (t-8)	0.0119	0.0768	0.0103	-0.0496	-0.0893***
	(0.0943)	(0.2469)	(0.0508)	(0.0344)	(0.0327)
D_1 Placebo (t-9)	-0.0432	-0.0373	-0.0324	-0.0233	-0.0049
	(0.0705)	(0.1887)	(0.0333)	(0.0258)	(0.0379)
D_1 Placebo (t-10)	-0.0027	0.1210	-0.0289	-0.0333	-0.0748***
	(0.0855)	(0.2030)	(0.0295)	(0.0305)	(0.0209)
D_2 Placebo (t-6)	0.1282	0.4583*	-0.0172	-0.0305	-0.0520
	(0.1022)	(0.2758)	(0.0425)	(0.0288)	(0.0433)
D_2 Placebo (t-7)	0.0068	0.7605***	0.0827**	-0.0466*	-0.0202
	(0.0832)	(0.1825)	(0.0359)	(0.0261)	(0.0373)
D_2 Placebo (t-8)	-0.0031	-0.2553	-0.0398	0.0114	0.0018
	(0.0730)	(0.2077)	(0.0274)	(0.0295)	(0.0353)
D_2 Placebo (t-9)	-0.0269	0.2823	0.0243	-0.0224	-0.0191
	(0.0649)	(0.2142)	(0.0255)	(0.0294)	(0.0322)
D_2 Placebo (t-10)	-0.0511	-0.4066**	-0.0141	-0.0095	0.0221
	(0.0590)	(0.2008)	(0.0347)	(0.0340)	(0.0335)
Observations	8169	8169	8169	8169	8169

Notes: OLS regression. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, the robust standard errors (in parentheses) are adjusted for clusters at the state × year of birth level. All models include state and year effects and state specific trends, a dummy for female, school starting age and state and year specific rules for the transition to upper secondary school (recommendation, binding grades, entrance examinations). The regression also includes leads from 0 to 5 and lags from 1 to 5.

which henceforward prohibited grade retention in the first years of school. In this case a placebo reform estimation would yield higher coefficients than estimates from the true reform.

Another threat to the identifying assumption might arise if federal states suddenly change quotas for upper secondary schools and increase the number of individuals transitioning to a higher track. This is possible, however, it is not relevant for the research question of interest which is finding out whether later grading in combination with socio-economic status has an effect on educational achievement. Also, there is a steady increase over the years of individuals moving to upper secondary school. The relatively small coefficient found for D_1 compared to the one for D_2 is also more in line with a causal interpretation of an effect of later grading since sudden changes in quotas should show a larger coefficient for D_1 as well as coefficients for D_1 and D_2 being more similar in size.

Socio-economic status As an alternative measure of socio-economic status the maximal parental occupational status is used. Three groups are constructed: high (OCC3: Managers and professional workers, small self-employed with employees), medium (OCC2: routine clerical work, routine service and sales work, small self-employed without employees, manual supervisors, skilled manual workers) and low (OCC1: semi- and unskilled manual workers, agricultural labor). The results can be seen in Table 2.8. For the transition to upper secondary school and the years of schooling none of the interaction coefficients are significant at the 10-percent level. This is in line with results for parental education. The same is true for the other outcome variables. Again, there is no clear pattern concerning the direction of the effects for parental occupation and there are mostly no significant differences between occupation groups.

2.7 Conclusion

In light of the literature of recent years that finds strong evidence of the effects of social origin on educational transitions, this study aims at finding out if replacing grades by verbal assessments in primary school can increase educational opportunity for children with low parental education. The reform analyzed meant that West German federal states abolished grades in school reports in the first classes of primary school starting

Table 2.8: DD estimates, educational outcomes on grading reform for parental occupation

	Transition to upp. sec. (1)	Years of schooling (2)	MS → US (3)	US → MS/LS (4)	Grade ret. 1-4 class (5)
Grading reform:					
D_1	-0.0117	-0.4931	0.0751	0.1030**	-0.0347
	(0.1042)	(0.3947)	(0.0546)	(0.0418)	(0.0516)
D_2	0.0193	-0.1719	0.0815	0.1083**	-0.0192
	(0.1048)	(0.3942)	(0.0590)	(0.0447)	(0.0514)
Parental Occupation (Reference: $OCC1$):					
$OCC2$	0.0976***	0.3577***	-0.0024	0.0098	-0.0375***
	(0.0171)	(0.0626)	(0.0097)	(0.0072)	(0.0099)
$OCC3$	0.3837***	1.3010***	0.0052	0.0312***	-0.0536***
	(0.0171)	(0.0673)	(0.0110)	(0.0079)	(0.0104)
Interaction of D_1 with $OCC1 - OCC3$ (Reference: $D_1 * OCC1$):					
$D_1 \times OCC2$	0.0097	0.0976	0.0427*	0.0221	0.0455**
	(0.0491)	(0.2386)	(0.0234)	(0.0220)	(0.0225)
$D_1 \times OCC3$	0.0119	0.0227	0.0199	0.0177	0.0251
	(0.0442)	(0.2052)	(0.0215)	(0.0212)	(0.0201)
Interaction of D_2 with $OCC1 - OCC3$ (Reference: $D_2 * OCC1$):					
$D_2 \times OCC2$	-0.0009	0.2647	0.0484**	-0.0014	-0.0052
	(0.0422)	(0.1774)	(0.0224)	(0.0182)	(0.0247)
$D_2 \times OCC3$	0.0590	0.1818	0.0269	-0.0127	-0.0374
	(0.0367)	(0.1666)	(0.0193)	(0.0177)	(0.0241)
Observations	7415	7415	7415	7415	7415

Notes: OLS regression. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, the robust standard errors (in parentheses) are adjusted for clusters at the state × year of birth level. All models include state and year effects and state specific trends, a dummy for female, school starting age and state and year specific rules for the transition to upper secondary school (recommendation, binding grades, entrance examinations).

in the 1970s. A difference-in-difference estimator is employed. This allows the use of variation given by the different timing of the implementation of the reform in federal states to identify a causal effect. The data comes from the National Educational Panel Study (NEPS) for Germany as well as law gazettes of federal states.

Results provide little evidence that the later grading reform had an effect on the likelihood of transitioning to upper secondary school as well as educational attainment. This means there is no evidence that not grading in the first two classes has an impact on track choice after class four. Furthermore, there is no evidence that students with different levels of parental education benefit differently from later grading which means educational opportunity seems to be unaffected by the reform. The reform does however increase transitions between secondary school tracks. This indicates an educational mismatch after primary school possibly caused by school reports without grades being a less precise signal of individual ability. There is, however, no selectivity regarding parental background for these transitions. The findings are robust to the inclusion of institutional rules concerning the transition to upper secondary school, other controls and using parental occupation as an indicator for socio-economic status.

In summary, the results suggest that it does not make a difference whether performance is judged by grades or verbal reports with regard to the subsequent track choice. Pekrun (1996) notes that other characteristics may be more important for the measurement of performance. These are characteristics such as the quality of the judgment, objectivity, to what extent judgments are displayed to individuals and the consequences of selection. While this study investigates the impact of not grading in the first two classes of primary school, future work may focus on finding out how a four year exposure to school reports without grades impacts educational opportunity. Also interesting would be to find out how other outcomes, for instance soft skills, are affected by not grading in primary school.

2.A Appendix

	BW	BY	BE	HB	HH	HE	NI	NW	RP	SL	SH
1954/1955	Y	Y	Y	Y	Y	Y	Y	L	Y	/	Y
1955/1956	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1956/1957	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1957/1958	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1958/1959	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1959/1960	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1960/1961	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1961/1962	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1962/1963	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1963/1964	Y	Y	N	Y	Y	Y	Y	Y	Y	/	Y
1964/1965	Y	Y	N	Y	Y	Y	Y	Y	Y	/	Y
1965/1966	Y	Y	N	Y	Y	Y	Y	Y	Y	/	Y
1966/1967	Y	Y	N	Y	Y	Y	Y	Y	Y	/	Y
1967/1968	Y	Y	N	Y	Y	Y	Y	Y	Y	/	Y
1968/1969	Y	Y	N	Y	N	Y	Y	Y	Y	/	Y
1969/1970	Y	Y	N	Y	N	Y	Y	Y	Y	/	Y
1970/1971	Y	Y	N	Y	N	Y	Y	Y	Y	/	Y
1971/1972	Y	Y	N	Y	N	Y	Y	Y	Y	/	N
1972/1973	Y	Y	N	Y	N	Y	Y	Y	Y	/	N
1973/1974	Y	Y	N	Y	N	Y	Y	Y	Y	/	N
1974/1975	Y	Y	N	Y	N	Y	Y	Y	Y	/	N
1975/1976	Y	Y	N	Y	N	Y	Y	Y	Y	Y	N
1976/1977	Y	Y	N	Y	N	Y	Y	Y	Y	Y	N
1977/1978	Y	Y	N	N	N	Y	Y	Y	Y	Y	N
1978/1979	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1979/1980	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1980/1981	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1981/1982	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1982/1983	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1983/1984	Y	Y	N	N	N	Y	N	Y	Y	Y	N
1984/1985	Y	Y	N	N	N	Y	N	Y	N	Y	N
1985/1986	Y	Y	N	N	N	Y	N	Y	N	Y	N
1986/1987	Y	Y	N	N	N	Y	N	Y	N	Y	N
1987/1988	Y	Y	N	N	N	Y	N	Y	N	Y	N
1988/1989	Y	Y	N	N	N	Y	N	Y	N	N	N
1990/1991	Y	Y	N	N	N	Y	N	Y	N	N	N
1991/1992	Y	Y	N	N	N	Y	N	Y	N	N	N
1992/1993	Y	Y	N	N	N	Y	N	Y	N	N	N
1993/1994	Y	Y	N	N	N	N	N	Y	N	N	N
1994/1995	Y	Y	N	N	N	N	N	Y	N	N	N
1995/1996	Y	Y	N	N	N	N	N	Y	N	N	N
1996/1997	Y	Y	N	N	N	N	N	Y	N	N	N

Notes: Secondary school recommendation is Y: binding, N: not binding. L: rules differ for local education authorities, /: rules missing. Adapted from Helbig and Nikolai (2015).

Table 2.A.1: Secondary school recommendation

	BW	BY	BE	HB	HH	HE	NI	NW	RP	SL	SH
1954/1955	N	N	N	N	N	N	/	L	N	/	N
1955/1956	N	N	N	N	N	N	N	L	N	/	N
1956/1957	N	N	N	N	N	N	N	L	N	/	N
1957/1958	N	N	N	Y	N	N	N	L	N	/	N
1958/1959	N	N	N	Y	N	N	N	L	N	/	N
1959/1960	N	N	N	N	N	N	N	L	N	/	N
1960/1961	N	N	N	N	N	N	N	L	N	/	N
1961/1962	N	N	N	N	N	N	N	L	N	/	N
1962/1963	N	N	N	N	N	N	N	L	N	/	N
1963/1964	N	N	N	N	N	N	N	N	N	/	N
1964/1965	N	N	N	N	N	N	N	N	N	/	N
1965/1966	N	N	N	N	N	N	N	N	N	/	N
1966/1967	N	N	N	N	N	N	N	N	Y	/	N
1967/1968	N	N	N	Y	N	N	N	N	Y	/	N
1968/1969	N	N	N	Y	N	N	N	N	Y	/	N
1969/1970	N	N	N	Y	N	N	N	N	Y	/	N
1970/1971	N	N	N	Y	N	N	N	N	Y	/	N
1971/1972	N	Y	N	Y	N	N	N	N	Y	/	N
1972/1973	N	Y	N	Y	N	N	N	N	Y	/	N
1973/1974	N	Y	N	Y	N	N	N	N	Y	/	N
1974/1975	N	Y	N	Y	N	N	N	N	Y	/	N
1975/1976	N	Y	N	Y	N	N	N	N	Y	/	N
1976/1977	N	Y	N	Y	N	N	N	N	Y	/	N
1977/1978	N	Y	N	N	N	N	N	N	Y	/	N
1978/1979	N	Y	N	N	N	N	N	N	Y	N	N
1979/1980	Y	Y	N	N	/	N	N	N	Y	N	N
1980/1981	Y	Y	N	N	Y	N	N	N	Y	N	Y
1981/1982	Y	Y	N	N	Y	N	N	N	Y	N	Y
1982/1983	Y	Y	N	N	Y	N	N	N	Y	N	Y
1983/1984	Y	Y	N	N	Y	N	N	N	Y	N	Y
1984/1985	Y	Y	N	N	Y	N	N	N	Y	N	Y
1985/1986	Y	Y	N	N	Y	N	N	N	Y	N	Y
1986/1987	Y	Y	N	N	Y	N	N	N	Y	N	Y
1987/1988	Y	Y	N	N	Y	N	N	N	Y	N	Y
1988/1989	Y	Y	N	N	Y	N	N	N	Y	N	Y
1990/1991	Y	Y	N	N	Y	N	N	N	Y	N	Y
1991/1992	Y	Y	N	N	Y	N	N	N	Y	N	N
1992/1993	Y	Y	N	N	Y	N	N	N	Y	N	N
1993/1994	Y	Y	N	N	Y	N	N	N	Y	N	N
1994/1995	Y	Y	N	N	Y	N	N	N	Y	N	N
1995/1996	Y	Y	N	N	Y	N	N	N	Y	N	N
1996/1997	Y	Y	N	N	Y	N	N	N	Y	N	N

Notes: Grades are binding or not binding for the transition to upper secondary school. Y: binding, N: not binding, L: rules differ for local education authorities, /: rules missing. Adapted from Helbig and Nikolai (2015).

Table 2.A.2: Criteria of transition: binding vs. non-binding grades

	BW	BY	BE	HB	HH	HE	NI	NW	RP	SL	SH
1949/1950	/	Y	S	Y	/	Y	/	L	/	/	Y
1950/1951	/	Y	S	Y	/	Y	/	L	/	/	Y
1951/1952	/	Y	S	Y	/	Y	/	L	Y	/	Y
1952/1953	/	Y	N	Y	/	Y	/	L	Y	/	Y
1953/1954	/	Y	N	Y	/	Y	/	L	Y	/	Y
1954/1955	Y	Y	N	Y	Y	Y	/	L	Y	/	Y
1955/1956	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1956/1957	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1957/1958	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1958/1959	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1959/1960	Y	Y	N	Y	Y	Y	Y	L	Y	/	Y
1960/1961	Y	Y	N	Y	Y	N	Y	L	N	/	Y
1961/1962	Y	Y	N	Y	Y	N	Y	L	N	/	Y
1962/1963	Y	Y	N	Y	Y	N	Y	L	N	/	Y
1963/1964	Y	Y	N	Y	Y	N	Y	Y	N	/	Y
1964/1965	Y	Y	N	Y	Y	N	N	N	N	/	Y
1965/1966	Y	Y	N	Y	Y	N	N	N	N	/	Y
1966/1967	Y	Y	N	Y	Y	N	N	N	N	/	Y
1967/1968	Y	Y	N	N	Y	N	N	N	N	/	Y
1968/1969	Y	Y	N	N	N	N	N	N	N	/	Y
1969/1970	Y	Y	N	N	N	N	N	N	N	/	Y
1970/1971	Y	Y	N	N	N	N	N	N	N	/	Y
1971/1972	Y	N	N	N	N	N	N	N	N	/	N
1972/1973	Y	N	N	N	N	N	N	N	N	/	N
1973/1974	Y	N	N	N	N	N	N	N	N	/	N
1974/1975	Y	N	N	N	N	N	N	N	N	/	N
1975/1976	Y	N	N	N	N	N	N	N	N	/	N
1976/1977	Y	N	N	N	N	N	N	N	N	/	N
1977/1978	Y	N	N	N	N	N	N	N	N	/	N
1978/1979	Y	N	N	N	N	N	N	N	N	N	N
1979/1980	N	N	N	N	N	N	N	N	N	N	N
1980/1981	N	N	N	N	N	N	N	N	N	N	N
1981/1982	N	N	N	N	N	N	N	N	N	N	N
1982/1983	N	N	N	N	N	N	N	N	N	N	N
1983/1984	N	N	N	N	N	N	N	N	N	N	N
1984/1985	N	N	N	N	N	N	N	N	N	N	N
1985/1986	N	N	N	N	N	N	N	N	N	N	N
1986/1987	N	N	N	N	N	N	N	N	N	N	N
1987/1988	N	N	N	N	N	N	N	N	N	N	N
1988/1989	N	N	N	N	N	N	N	N	N	N	N
1990/1991	N	N	N	N	N	N	N	N	N	N	N
1991/1992	N	N	N	N	N	N	N	N	N	N	N
1992/1993	N	N	N	N	N	N	N	N	N	N	N
1993/1994	N	N	N	N	N	N	N	N	N	N	N
1994/1995	N	N	N	N	N	N	N	N	N	N	N
1995/1996	N	N	N	N	N	N	N	N	N	N	N
1996/1997	N	N	N	N	N	N	N	N	N	N	N

Notes: Entrance examination for the transition to upper secondary school. Y: entrance examination is exclusive criteria or one of many criteria, N: no entrance examination, S: single school type, L: rules differ for local education authorities, /: rules missing. Adapted from Helbig and Nikolai (2015).

Table 2.A.3: Criteria of transition: entrance examination

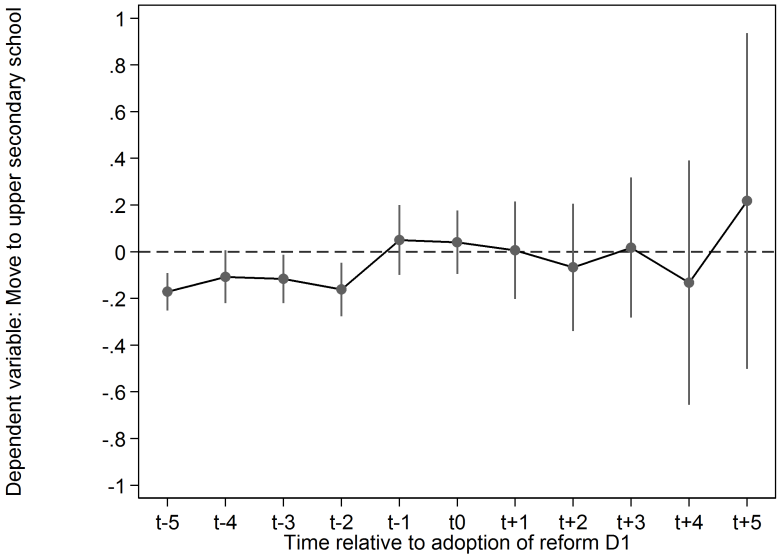


Figure 2.A.1: Estimates for D_1 from Equation 2.3 ($\alpha_{+5}, \alpha_{+4}, ..., \alpha_{-4}, \alpha_{-5}$), vertical lines mark 90% confidence interval

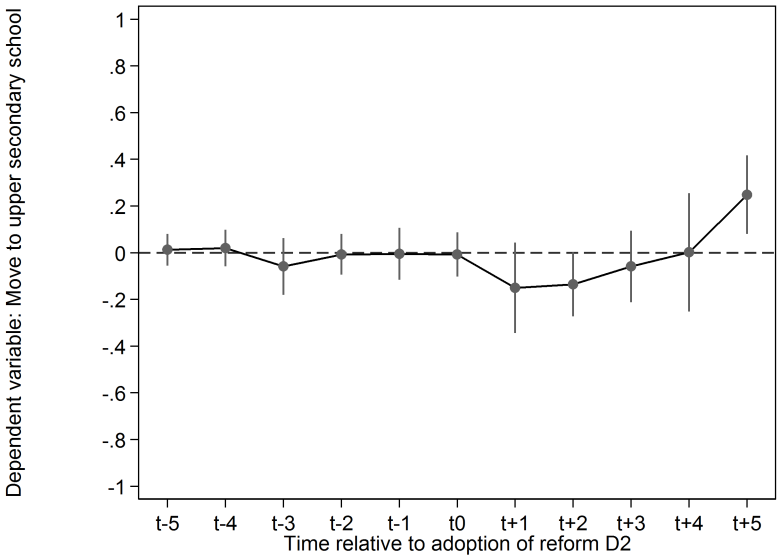


Figure 2.A.2: Estimates for D_2 from Equation 2.3 ($\beta_{+5}, \beta_{+4}, ..., \beta_{-4}, \beta_{-5}$), vertical lines mark 90% confidence interval

Chapter 3

The Long-term Effect of Age at School Entry on Competencies in Adulthood*

3.1 Introduction

Many countries determine the age at which children may legally start school by defining cut-off dates for enrollment. All children born before the cut-off are supposed to enter school in a given year, while those born after the cut-off are expected to wait until the start of the next school year. This leads to considerable between-child variation in the school starting age (SSA) within a class. A vast empirical literature shows that SSA has important effects on children's school performances. Children that are enrolled at a higher SSA outperform their younger classmates in mathematics, reading, and writing (see, e.g., Bedard and Dhuey, 2006; Elder and Lubotsky, 2009; Smith, 2009). Most countries also give parents legal options to enroll their children with a one-year delay. These options usually allow parents of children, who would have had a low SSA

*This chapter is based on joint work with Katja Görlitz (FU Berlin, DIW, RWI and IZA) and Marcus Tamm (RWI and IZA). This chapter uses data from the National Educational Panel Study (NEPS): Starting Cohort Adults, doi:10.5157/NEPS:SC6:8.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Katja Görlitz and Marcus Tamm acknowledge financial support from the Deutsche Forschungsgemeinschaft (GO 2168/1-2 and TA 829/2-2).

if they were enrolled regularly, to delay school entry by one year, making the child a high SSA student. This practice, which is often called "red-shirting" in the academic literature, has become increasingly common in recent years in the US and Germany (Deming and Dynarski, 2008; Statistisches Bundesamt, 2017, 2018). Parents might consider red-shirting to be an optimal decision if enrolling their children with a lower SSA is accompanied by lifetime disadvantages for their children. However, previous literature, which documents the negative SSA effects, focuses mostly on the test score differentials of schoolchildren. Only a handful of papers go beyond Grade 10 and show that SSA effects become much smaller when looking at IQ and SAT scores at around the age of 18 (see, e.g., Black et al., 2011; Hurwitz et al., 2015).

This paper is the first to investigate whether SSA test score differentials continue into adulthood or fade away after leaving school. It evaluates SSA effects on adult competencies, as measured in comprehensive tests administered as part of a representative survey of individuals between 23 and 71 years old. Analyzing the persistence of SSA effects provides important information for parents. It also contributes to the literature investigating the long-run effects of SSA on individuals' wages and employment (e.g., Black et al., 2011; Fredriksson and Öckert, 2014; Larsen and Solli, 2017). Individuals with a lower SSA have higher wages and better employment perspectives shortly after entering the labor market. This result is due to a lower SSA being accompanied by longer tenure and actual experience when holding the age constant. However, Fredriksson and Öckert (2014) and Larsen and Solli (2017) conclude that SSA has no effects on cumulative earnings over an individual's life using data for Sweden and Norway, respectively. Since one of the potential channels for long-run effects on wages and employment is SSA-induced differences in adult competencies, our study should be viewed as complementary to this stream of literature.[1]

Our study also contributes to the literature showing that SSA test score differentials decrease as children progress through school (see, e.g., Bedard and Dhuey, 2006; Elder and Lubotsky, 2009). Even though many studies support this view, there is still "some disagreement about whether the effects are attenuated by middle school" (Cook and

[1]Other papers on the long-run effects of SSA are only slightly connected to our research question. These papers involve studies investigating the effects of SSA on crime (e.g., Cook and Kang, 2016; Landersø et al., 2017), fertility (e.g., McCrary and Royer, 2011; Skirbekk et al., 2004), and marriage outcomes (e.g., Lefgren and McIntyre, 2006).

Kang, 2018, p. 2). Apart from our main analysis, which sheds further light on the potential attenuation by presenting the long-term effects of SSA, our literature section makes an additional contribution. It graphically analyzes how the estimates of SSA on students' test scores presented in the previous literature differ by grade level.

The identification of the effect of SSA on competencies relies on an instrumental variable strategy that exploits the state- and year-specific rules given by the cut-off dates. The empirical model controls for a full set of month-of-birth dummies. This is important because Buckles and Hungerman (2013) point out that the distribution of family background differs by children's months of birth and differences in SSA might capture some of these effects. This capture is possible in our study because the school-entry regulations were subject to several changes at the state and year levels, allowing separate identification of the month-of-birth effects. In addition, our model can account for "age-at-test effects". Such an accounting is not possible in most of the previous studies that analyze SSA effects on test scores for school-aged children. If all children take the test at the same point in time, being older at school entry means automatically taking each test at an older age. Black et al. (2011) show that SSA effects become much smaller after controlling for the age at the time of testing. We are able to account for the age at the time of testing because the competency tests were not taken at a particular date for all individuals, but the time interval for the interviews stretched out over several months and the test-taking date was unrelated to the date of birth and the state where individuals went to school.

This paper contributes to the literature by additionally disentangling relative from absolute age effects, which is important from a policy perspective. Relative age measures the age difference compared to the ages of the other students within the cohort. Absolute age refers to the age (and, thus, maturity) when starting school. The previous literature is generally not able to separate these two potential channels when analyzing SSA differences. This inability to separate comes about due to the fact that the relative age at school entry is linearly related to the absolute age. One exception is the work of Cascio and Schanzenbach (2016), which separates absolute from relative SSA effects. They show that absolute age significantly increases a combined math and reading test score in Grade 8, while relative age has a statistically negative impact on the test score. Our analysis provides estimates for relative and absolute SSA effects by exploiting the fact that several states experienced changes in cut-off dates over time. However, to

compare our results to the previous literature, our baseline results provide SSA effects without separating the two effects.

Our analysis shows that the impact of SSA on math and text comprehension measured in adulthood are considerably smaller than what the literature has shown for children in school. Further, both estimates are statistically insignificant. In contrast, the effect of SSA on receptive vocabulary is sizable in adulthood, with a one-year-higher SSA increasing competency by around a third of a standard deviation. These findings survive several tests of robustness. When disentangling the effect of SSA into an absolute and a relative age effect, we find that receptive vocabulary is affected solely by absolute age.

The remainder of the paper is organized as follows. Section 3.2 gives an overview of the previous literature on SSA effects on test scores and illustrates how the estimated effects differ by grade level. Section 3.3 describes the data and the school entry regulations, and Section 3.4 presents the estimation strategy. The fifth section presents and discusses the results, and the final section offers a conclusion.

3.2 Previous literature

There exists a vast empirical literature that estimates the causal effects of SSA on test scores. The majority of papers rely on instrumental variable strategies for identification. The instruments exploit the variation in individuals' dates of birth and cut-off rules. Both determine at which age a student should legally start school. Alternatively, some studies use the information on date of birth and the cut-off rules for implementing a regression discontinuity design. The previous literature uses test scores measured at different grade levels as outcomes. The earliest test score differences are measured during kindergarten, where the variation derives from kindergarten entry cut-off dates. Several studies show that the oldest kindergarten children score significantly better in reading and math tests compared to the youngest children within the class (Datar, 2006; Elder and Lubotsky, 2009; Lubotsky and Kaestner, 2016; Cascio and Schanzenbach, 2016). Being one year older within a kindergarten class increases math and reading/writing test scores by 0.43 to 0.87 and 0.42 to 0.58 of a standard deviation, respectively.

These early advantages of being the oldest when entering school continue into higher classes (Bedard and Dhuey, 2006; Fredriksson and Öckert, 2006; Puhani and Weber,

2007; McEwan and Shapiro, 2008; Ponzo and Scoppa, 2014; Cook and Kang, 2016; Dhuey et al., 2017; Attar and Cohen-Zada, 2018; Koppensteiner, 2018). However, when comparing the magnitude of the relationship, this advantage seems to decrease when children progress through school. In Grades 9 and 10, the test score differential from being one year older at school entry is 0.10 to 0.20 of a standard deviation in math and 0.15 to 0.24 in reading/writing (Smith, 2009; Black et al., 2011; Peña, 2017).

We provide further evidence on how SSA test score differential evolves by graphically illustrating how SSA estimates presented in the literature differ by grade level. The studies considered for the graphs come from an extensive literature search performed using EconLit.[2] Figure 3.1 shows the relationship for math and reading/writing scores in Panels A and B, respectively. It illustrates that the test score advantage conferred on the oldest students decreases with grade level. It is largest in kindergarten – even more so for math compared to reading/writing. The dashed lines indicate the best fit for the functional relationship between SSA test score differentials and grade levels.[3] The lines suggest a declining but non-linear relationship. Based on the graphs, extrapolation beyond Grade 10 suggests that the math test score differentials will continue to decline, while there is no clear conclusion for reading/writing.

There are only a handful of papers that go beyond Grade 10. However, most of these studies analyze competencies at around the age of 18 and do not extend to higher ages. Further, they sometimes use tests that do not represent competencies but, rather,

[2]The search on EconLit was done on November 20th, 2018. It included the following keywords: "school entry age", "kindergarten entry age", "school entrance age", "kindergarten entrance age", "school starting age", "kindergarten starting age", "age at school entry", "age at kindergarten entry", "age at school start", "age at kindergarten start", "enrollment cutoff", "age effect school", "relative age school", and "relative age performance". From the studies found in this manner, we kept only those in which the identification strategy exploits cut-off rules and the dependent variable is a test score (rather than grades given by the teacher). To be able to consider all estimates from these studies in one figure for math and in one figure for reading/writing, the estimates additionally had to fulfil the following criteria: i) they are provided separately for math and reading/writing tests; ii) they are provided separately by grade level; iii) they can be interpreted in terms of standard deviations of the test score's distribution (i.e., summary statistics that at least allowed a corresponding interpretation had to be provided); and iv) the estimates are provided for the entire population of students rather than just separately for subgroups, such as boys and girls. Table 3.A.1 in the appendix contains the full list of studies that are used for the figures.

[3]To decide on the functional form of the relationship, we have run linear, quadratic, cubic, and exponential regressions. According to the AIC criteria, a cubic specification best fits the relationship for math scores, and a quadratic model works best for reading/writing.

abilities, or they present their results for a selective sample of students. Therefore, we could not include these results in Figure 3.1. The following paragraph will summarize these findings.

Using cut-off dates for identification, Black et al. (2011) provide evidence of the impact of SSA on men's IQ scores measured at 18 years of age. They find relatively small negative effects, i.e., those who are older within class have slightly lower IQ scores. When using IQ scores as an outcome, one caveat is that it is unclear to what extent IQ represents competencies acquired at school and to what extent it represents innate abilities. Cascio and Lewis (2006) provide reduced form estimates of the effect of SSA on AFQT scores of individuals aged 15 to 19. While the effects are not significant for whites, they are slightly negative for blacks. Fletcher and Kim (2016) estimate the reduced form effects of the kindergarten entry cut-offs that differ by state on state-specific averages of the test scores in math and reading. They find no effect of kindergarten entry age on test scores in Grade 12. Nam (2014) exploits cut-off rules using Korean data. The results indicate that SSA has positive effects on math and reading/writing test scores in Grades 6 to 8 but that these differences do not persist when students graduate from high school. Implementing a regression discontinuity design based on cut-off dates, Hurwitz et al. (2015) show that SSA has no significant impact on the SAT scores of college-bound students. However, both Nam (2014) and Hurwitz et al. (2015) analyze test scores of students who voluntarily participated in the test, thus representing a selective sample of students who intend to enroll in college after high school graduation. The same is true for Pellizzari and Billari (2012), who look at the performance of university students. They find that younger students outperform their older fellow students, which they explain via fewer social activities and, thus, more learning time for the younger students. In conclusion, it is an open question whether and how the effects of SSA on competencies linger into adulthood.

3.3 Data and school enrollment regulations

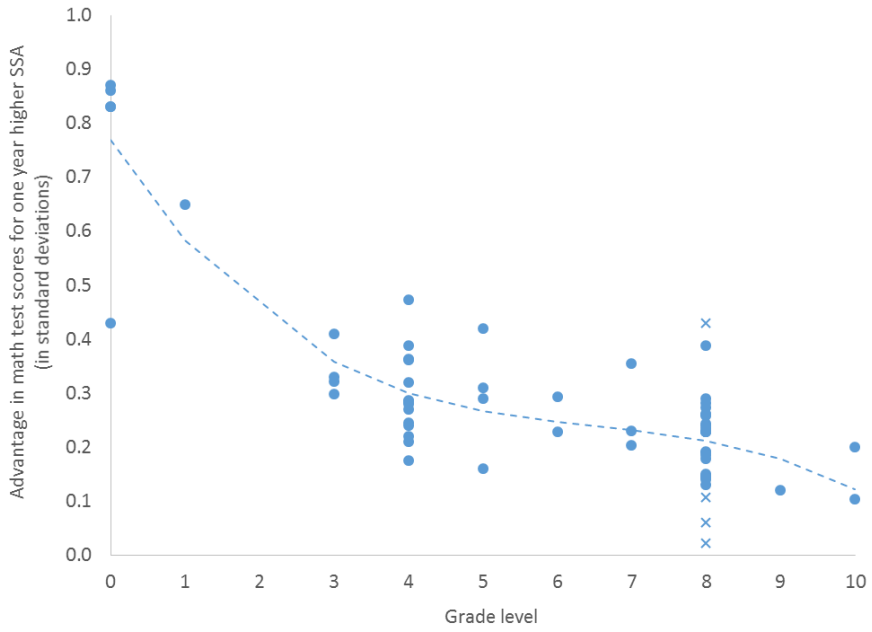
3.3.1 Data

The analysis is based on the adult cohort of the National Educational Panel Study (NEPS-SC6). The NEPS-SC6 includes information on the educational, occupational, and family formation processes for individuals born between 1944 and 1986. It covers detailed information from birth through adult life (Blossfeld et al., 2011), including information on state and date of school entry. This last information allows us to determine accurately the date when children should have entered school according to official regulations and when they have actually done so.

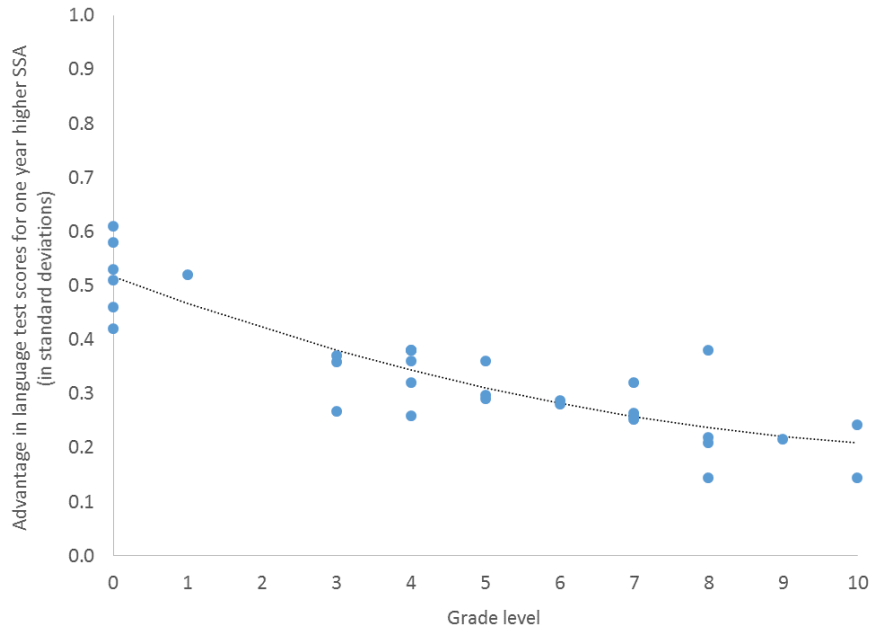
The data also contain information on competencies measured in adulthood. For our analysis, we use one test for mathematical competencies and two tests for language competencies, i.e., for text comprehension and receptive vocabulary. The different competency tests were collected in the NEPS-SC6 data in different waves. The tests were part of the interviews in wave 2010/2011, in which individuals were randomly selected to take the math or the text comprehension test or both, and 2012/2013, in which text comprehension tests were administered to individuals who did not take the test in 2010/2011. In 2014/2015, tests for receptive vocabulary were part of the interview for all respondents. These competency tests capture basic competencies in everyday life situations (Weinert et al., 2011). The math competency tests were designed to describe respondents' abilities to use and apply mathematics flexibly in realistic situations. They cover four content areas: data and probability, quantities, shape and space, and change and relationship (Schnittjer and Durchhardt, 2015). To provide one example from the area data and probability, respondents are asked whether they understand the statistics on side effects from the package inserts of a pharmaceutical product. The test for text comprehension uses different types of text from which respondents must find information in the text, draw text-related conclusions, and reflect and assess (Gehrer et al., 2012). The receptive vocabulary test is similar to the Peabody Picture Vocabulary Test. Respondents have to assign pictures to a single word given by the interviewer by choosing from four possibilities. The correlation between the three test scores is high, with correlation coefficients varying between 0.47 and 0.55, but there is substantial independent variation in each of these outcomes, i.e., we are guaranteed that the three tests scores measure different dimensions of competencies. In order to ease interpretation of

Figure 3.1: Effects of school starting age on test scores by grade level

Panel A: Math test scores



Panel B: Reading/writing test scores



Notes: Grade 0 refers to kindergarten. The dots represent statistically significant estimates and the crosses statistically insignificant estimates. The dashed blue line presents the best fit of a model, where grade level enters as a third polynomial. The dotted black line presents the best fit of a quadratic relationship between reading test scores and grade level. The AIC criteria was used to decide on the functional form of the relationships. Table 3.A.1 in the appendix shows the full list of studies considered in the figures.

the results, the test scores are normalized to having a mean of zero and a variance of one (z-scores).

Figure 3.A.1 in the appendix illustrates the distribution of the three test scores. Even though the tests are designed to capture basic competencies, there is ample variation at the upper and lower end of the competency distributions. The histogram of the receptive vocabulary test score shows that its distribution is highly left-skewed. Therefore, the results section also includes findings for receptive vocabulary that analyze whether this skewness presents a problem. Specifically, we transform the receptive vocabulary test scores using a Box-Cox-transformation. The transformed and standardized test scores have a distribution that is much more similar to a normal distribution (see Figure 3.A.2 in the appendix). Because the interpretation of the transformed test scores is less straightforward than that for the untransformed test scores, the main specification for receptive vocabulary is based on the untransformed test scores. However, the robustness section documents that the main results remain unchanged when using the transformed test scores.

The analysis focuses on individuals who entered primary school in West Germany. East Germany (including Berlin) is dropped from the analysis because the East German schooling system differed considerably from the schooling system in the West for this time period. Also, the East German cut-off dates for school entry did not differ between regions and over time, inhibiting separate identification of the effects of age at school entry from month-of-birth effects. The analysis considers information on around 3,700 individuals for mathematical literacy, around 5,900 individuals for text comprehension, and around 6,000 individuals for receptive vocabulary.[4] Descriptive statistics of the three samples are provided in Table 3.1. Differences between the samples are small, except for age at test, which is due to the fact that the different competencies were assessed in different waves. The average age at test is between 46 and 50 years. Half of the respondents are women. On average, they entered school at $6\frac{1}{2}$ years old and received an average of $13\frac{1}{2}$ years of education.

[4]The initial West German sample size of individuals participating in the competency tests is 3,766, 5,976, and 6,190, respectively. We had to drop between 1.4% and 1.7% of the observations to ensure that all variables required to construct the instrument are observable and due to data cleaning.

Table 3.1: Descriptive statistics

	Sample: Mathematical literacy		Sample: Text comprehension		Sample: Receptive vocabulary	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.499	0.500	0.498	0.500	0.503	0.500
Age at school entry	6.48	0.49	6.48	0.49	6.48	0.49
Age at test	45.80	11.19	46.68	11.38	49.61	11.09
Years of education	13.58	2.70	13.39	2.79	13.54	2.76
Observations	3,678		5,855		6,053	

Notes: The reason for the differences in the number of observations by competency test is that the tests were taken in different waves.

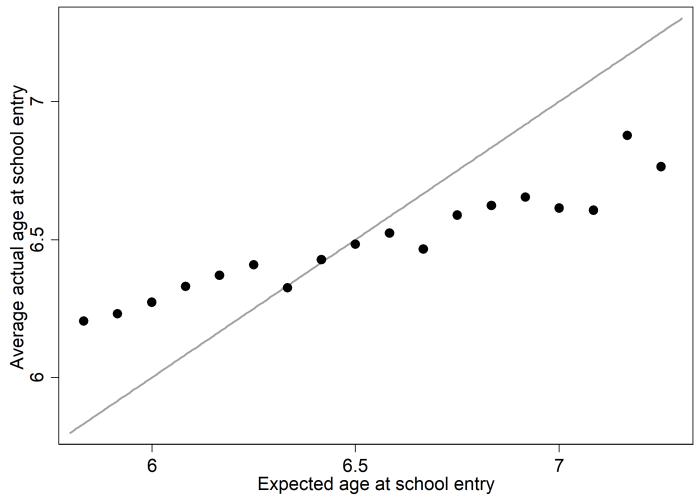
3.3.2 School enrollment regulations

The analysis exploits variation in school entry cut-off dates that legally define at which age children are supposed to enter school. In Germany, children turning age 6 before the cut-off date enter school at the beginning of the school year, while those turning 6 after the cut-off enter school one year later. The cut-off date is determined by the states. The age at school entry does not depend only on the cut-off date and one's birthday; it also depends on the date when the school year starts, which is also legally determined at the state level. The cut-off date and the beginning of the school year have experienced several changes over time in German states. In recent decades, the school year has generally started in August, while it started in April during the 1950s and 1960s (see Table 3.A.2 in the appendix). The cut-off date experienced more changes, varying from the 31st of March to the 31st of December (see Table 3.A.3 in the appendix).

The cut-off is not strictly enforced, as there are options to enroll earlier or to delay school entry, but a sizeable share of parents sticks to the official cut-off. Figure 3.2 documents how the legally determined expected school starting age is related to the average of the actual SSA. The diagonal line illustrates the case of full compliance, i.e., when all children follow the cut-off rules without exception. The dots show the actual association between expected and average actual school starting age. Even though Figure 3.2 documents a clear, positive association between the expected and the actual relationship, it also shows that the actual relationship is flatter. This flatness indicates

that children with relatively low expected entry ages, i.e., turning 6 years old just before the school cut-off, are, on average, older than 6 when entering school since some children delay entry. Further, children with relatively high expected entry ages, i.e., those turning 7 shortly after the cut-off, are, on average, somewhat younger than 7 when entering school because some children enroll early. This pattern is evidence that parents actually make use of the legal possibilities of delayed and early school entry. Since non-compliance with the cut-off rules occurs and might be selective, exploiting variation from legal cut-off dates to identify the causal effect of age at school entry seems important. Also, note that the range of expected age at school entry in Figure 3.2 exceeds one year. This is because the minimum expected age at school entry differs between the states in Germany. While some states have official rules that students should be enrolled between the ages of 5.8 and 6.7 (when complying with the rules), others allow the age range to vary from 6.4 to 7.3 years. This variation is exploited in a further analysis to find out whether the estimated differences in competencies by the age at school entry derives from differences in absolute or relative age.

Figure 3.2: The relationship between the expected and the average of the actual age at school entry



Note: The dots indicate the association between the expected school starting age and the average actual age at school entry. The diagonal line indicates how the association would look if everyone complied with the cut-off rules.

3.4 Estimation strategy

The causal effect of school starting age is estimated by the following equation:

$$Y_i^k = \beta_0^k + \beta_1^k SSA_i + X_i\delta^k + \varepsilon_i^k, \tag{3.1}$$

where Y_i^k is the standardized test score of individual i on test k (k = mathematical literacy, text comprehension, or receptive vocabulary). SSA_i is school starting age, and X_i is a set of covariates, including gender, dummies for the states of enrollment in primary school, a full set of year-of-birth dummies, and a full set of month-of-birth dummies. Estimating the parameter of interest β_1^k from Equation 3.1 by OLS would induce biased estimates because the early or late enrollment of children varies with children's abilities and parents' resources in a systematic manner. For instance, Dobkin and Ferreira (2010) show that children of highly educated parents have a lower probability of complying with the enrollment regulations. These identification issues are taken into account by using an instrumental variable (IV) estimator. The instrument is based on the school entry cut-off dates that vary by states and over time. Following the previous literature, the instrument is defined as the expected age at school entry $ESSA_i$. This is the age at which an individual would have entered school if school entry were determined solely by the official regulations. The IV estimator is implemented in the following first-stage equation:

$$SSA_i = \tilde{\gamma}_0 + \tilde{\pi}' ESSA_i + X_i\tilde{\theta} + \tilde{\mu}_i. \tag{3.2}$$

X_i includes the aforementioned set of covariates. The inclusion of month-of-birth dummies is only possible because the cut-off dates vary over time and between states (see Section 3.3.2). Few papers can account for month-of-birth effects, although Buckles and Hungerman (2013) find that parents' backgrounds correlate with children's month of birth, even in two consecutive months within the same season. This correlation might lead to systematic differences between those born before and those born after the cut-off. Using only cut-off dates without controlling for the month of birth absorbs these differences, biasing the estimation of the effect of school starting age. Including the year-of-birth dummies serves the purpose of controlling for changes over time that likely correlate with the outcome and the instrument, such as the educational expansion

during the 1960s and 1970s. The year-of-birth dummies also capture the influence of age at test in a non-parametric way.[5]

All estimates account for clustering at the level of states. Since only ten clusters can be accounted for, as there are only ten states, the standard errors may suffer from downward bias. We follow suggestions in Cameron et al. (2008) for estimation with few clusters and present small-sample-adjusted p-values of a test against zero instead of the downward biased standard errors. As a test of robustness, we also present p-values of a wild cluster bootstrap, which was found to work well when the number of clusters is small (Cameron et al., 2008).

First-stage estimates and validity of the identification strategy

Table 3.2 presents the first-stage estimates. Increasing the expected age at school entry by one year is associated with an average increase of actual age at school entry by 0.38 years. Since the F-test for the significance of the instrument is always considerably above 10, there is no problem of a weak instrument. These results reinforce the conclusions from Figure 3.2, meaning that the expected school starting age is well suited to serve as an instrument for actual school starting age.

Due to the alternative specifications of the dependent variable being exactly identified, the validity of the instrument cannot be tested. However, in the following we perform several of measures to provide some plausibility arguments for the validity of the identification strategy. It has been suggested to analyze whether births are systematically displaced around the cut-off (McCrary, 2008). Figure 3.A.3 in the appendix shows the number of observations by distance to the cut-off for each of the three samples. No systematic pattern of bringing forward or postponement of births becomes visible.[6] In addition to this we also show that predetermined variables, such as parental characteristics, are not correlated with the instrument. If the instrument were correlated with parental background, this could have a direct effect on competencies. Table 3.3 shows, separately for the three samples, that the family status at the age of 15 (i.e., whether the individual was raised by a single parent) and mother's and father's education, age at birth, and migration status are all unrelated to the instrument. This suggests that

[5]We apply further robustness tests that include, directly, age at test measured on a monthly basis in the regression. The results remain the same.

[6]For an example of systematic timing of the date of delivery with respect to a policy reform in Germany, see Tamm (2013).

Table 3.2: The effect of the instrument on school starting age (first stage)

	Sample: Mathematical literacy	Sample: Text comprehension	Sample: Receptive vocabulary
Expected school starting age (ESSA)	0.3755	0.3697	0.3761
	(0.000)	(0.000)	(0.000)
Control variables	Yes	Yes	Yes
F-statistic (excluded instrument)	81.25	133.29	148.19
Observations	3,678	5,855	6,053

Note: The dependent variable is the school starting age. The first row indicates the sample used to analyze each of the three competencies. As mentioned in Section 3.3.1, using different samples is necessary because each competency was tested in a different survey wave. The control variables include the year of birth, the month of birth, the state of primary school enrollment, and gender. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

parents do not strategically plan to deliver children before or after the school cut-off date in Germany, which is also confirmed by Bahrs and Schumann (2016). In order to test the robustness of our main results, Section 3.5 also presents IV estimates controlling for the above-mentioned parental characteristics.

3.5 Results

Table 3.4 presents the main IV estimates in Panel A. The first column of Table 3.4 presents the results for mathematical literacy. The estimate is statistically insignificant and small in size, suggesting that SSA has no long-lasting effects on mathematical competencies that persist until adulthood. The point estimate for an increase of age at school entry by one year relates to 6% of a standard deviation, which is far below what was measured in Grade 10 (see Panel A of Figure 3.1). The second column of Table 3.4 documents the results using text comprehension as an outcome. Similar to the results for mathematical literacy, the estimate for text comprehension presented in Panel A is statistically insignificant, and it is much smaller than the effects that were found in the literature for tests taken at younger ages. Specifically, the point estimate relates to 8% of a standard deviation. The third column of Table 3.4 shows results using receptive

Table 3.3: The effect of school starting age on predetermined characteristics (IV results)

	Single-parent family at age 15	Mother with college degree	Father with college degree	Mother's age at birth	Father's age at birth	Mother foreign born	Father foreign born
Sample: mathematical literacy							
IV estimate: SSA	-0.0284	0.0056	0.0383	1.8301	2.0310	-0.0156	-0.0194
	(0.436)	(0.854)	(0.526)	(0.260)	(0.253)	(0.733)	(0.562)
Observations	3,674	3,586	3,553	3,573	3,535	3,657	3,606
Sample: text comprehension							
IV estimate: SSA	0.0692	0.0095	0.0605	0.3586	0.9372	-0.0129	-0.0404
	(0.150)	(0.561)	(0.223)	(0.433)	(0.444)	(0.729)	(0.291)
Observations	5,848	5,674	5,627	5,687	5,627	5,823	5,736
Sample: receptive vocabulary							
IV estimate: SSA	0.0482	0.0220	0.0976	0.2359	1.4198	-0.0435	0.0075
	(0.129)	(0.341)	(0.161)	(0.740)	(0.287)	(0.093)	(0.734)
Observations	6,045	5,870	5,819	5,884	5,813	6,018	5,933

Note: The table provides IV estimates of the effect of school starting age on the outcomes listed in the first row. ESSA is used as the instrument. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

vocabulary as an outcome. In contrast to the other two measures of competency, the impact of the estimate of school starting age on receptive vocabulary is significantly positive. An increase of age at school entry by one year is associated with an increase in test scores by a third of a standard deviation. This is an economically sizable effect.

Before discussing reasons for why the effect of SSA on the Peabody Picture test persists, while it vanishes for math skills and text comprehension, we present several robustness checks. Panel B of Table 3.4 shows that the results remain robust when controlling for parental background characteristics. This is not surprising given that Table 3.3 has already shown that parental background is uncorrelated with the instrument. We also find that controlling for parental background increases the precision of the estimates. This leads to lower p-values for the estimate of school starting age on receptive vocabulary but leaves our conclusion of insignificant results for math skills and text comprehension unchanged. As has already been discussed in Section 3.2, some recent studies have shown that controlling for age at test matters considerably when estimating the effects of SSA on competency tests (see, e.g., Black et al., 2011). In our

main specification, age at test is partially controlled for by the year-of-birth dummies. Panels C and D directly control for age at test measured on a monthly basis to capture even small differences in age at test. Small age differences within a given cohort occur because the interviews were conducted in different months over the year and because some individuals participated in the text comprehension test in wave 2010/2011, while others participated in wave 2012/2013. Panels C and D present the results of a linear specification for age at test and of a squared specification, respectively. Both specifications still control for the year-of-birth dummies. Results in Panels C and D are very similar to those of the main specification.

States that implement reforms to the schooling system, such as changes in the cut-off date, might experience different trends in the outcome variable even without reforms. To rule out that such state-specific changes over time affect our estimates, Panel E controls for state-specific time trends using a linear specification, and Panel F controls for state-specific time trends using a squared specification. Controlling for time trends has only a minor impact on the estimates of SSA. Panel G of Table 3.4 shows results excluding the most recent cohorts. We do so because most of the changes in cut-off dates and the school start month took place during the 1950s and 1960s. In order to not rely on cohorts that are relatively far away from these reforms, the specification in Panel G is restricted to cohorts born between 1944 and 1973, dropping those born between 1974 and 1986.[7] The results for this reduced sample are very similar to those for the entire sample.

Panels H and I present results for alternative definitions of the instrument. While Panel A includes a linear specification for ESSA as an instrument for SSA, Panel H additionally includes ESSA-squared as an instrument. Furthermore, Panel I uses separate dummies for each value of ESSA as an instrument for SSA. These alternative definitions of the instrument generally confirm our main findings of statistically significantly effects of SSA on receptive vocabulary, although the point estimate is reduced by around one third (from 0.35 to 0.21) in Panel I.

Panel J shows that the results are robust to calculating the standard errors by alternative methods. As is pointed out in Section 3.4, all estimates account for clustering at the level of the states. Because the number of clusters is small, we follow Cameron

[7] The last cohort affected by reforms is the birth cohort of 1963. 1973 is chosen to include, at most, ten birth cohorts afterwards.

Table 3.4: The effect of school starting age on competencies in adulthood

	Mathematical literacy (1)	Text comprehension (2)	Receptive vocabulary (3)
Panel A Main specification			
IV estimate: SSA	0.0648	0.0818	0.3461
	(0.795)	(0.644)	(0.021)
F-statistic (excluded instrument)	81.25	133.29	148.19
Observations	3,678	5,855	6,053
Panel B Controlling for parental background characteristics			
IV estimate: SSA	0.0492	0.0994	0.3797
	(0.839)	(0.573)	(0.009)
F-statistic (excluded instrument)	78.99	121.89	126.57
Observations	3,678	5,855	6,053
Panel C Controlling for age at test in addition to the year-of-birth dummies			
IV estimate: SSA	0.0628	0.0849	0.3462
	(0.800)	(0.649)	(0.021)
F-statistic (excluded instrument)	81.30	132.70	148.14
Observations	3,678	5,855	6,053
Panel D Controlling for age at test and its square in addition to the year-of-birth dummies			
IV estimate: SSA	0.0422	0.0864	0.3668
	(0.873)	(0.653)	(0.028)
F-statistic (excluded instrument)	94.73	205.61	171.34
Observations	3,678	5,855	6,053
Panel E Controlling for state-specific time trends			
IV estimate: SSA	0.0712	0.0765	0.3141
	(0.767)	(0.643)	(0.043)
F-statistic (excluded instrument)	80.18	132.95	155.91
Observations	3,678	5,855	6,053
Panel F Controlling for state-specific time trends using a quadratic specification			
IV estimate: SSA	0.0742	0.0809	0.2710
	(0.771)	(0.612)	(0.075)
F-statistic (excluded instrument)	77.14	131.06	153.7
Observations	3,678	5,855	6,053
Panel G Sample 1944-1973			
IV estimate: SSA	0.1000	0.1015	0.2907
	(0.707)	(0.584)	(0.015)
F-statistic (excluded instrument)	71.61	106.66	127.47
Observations	3,028	4,799	4,984
Panel H Alternative definition of the instrument: ESSA + ESSA-squared (as additional instrument)			
IV estimate: SSA	0.061	0.0497	0.3091
	(0.809)	(0.772)	(0.047)
F-statistic (excluded instruments)	40.74	69.20	77.21
Observations	3,678	5,855	6,053
Panel I Alternative definition of the instrument: Separate dummies for each value of ESSA			
IV estimate: SSA	-0.0498	-0.0122	0.2110
	(0.803)	(0.934)	(0.054)
F-statistic (excluded instruments)	7.70	80.35	131.60
Observations	3,678	5,855	6,053
Panel J Wild cluster bootstrap			
IV estimate: SSA	0.0648	0.0818	0.3461
	[0.904]	[0.667]	[0.096]
F-statistic (excluded instrument)	81.25	133.29	148.19
Observations	3,678	5,855	6,053

Note: The table presents results of IV regressions that use ESSA as an instrument for SSA (see equations 1 and 2). The outcome variables are documented in the first row. Panel B controls for the parental background characteristics that are shown in Table 3. Standard errors account for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses. Panel J presents p-values using wild cluster bootstrap that are shown in brackets.

et al. (2008) by presenting small-sample-adjusted p-values in our main specification. An alternative procedure for obtaining inference is using wild cluster bootstrapping, which is implemented in Panel J. The conclusion remains unchanged, i.e. the IV estimate for receptive vocabulary is statistically significant, while the estimates for mathematical literacy and for text comprehension remain statistically insignificant.

Table 3.5: Robustness check using the transformed receptive vocabulary test score

	Language: transformed receptive vocabulary
IV estimate: SSA	0.3496
	(0.034)
F-statistic (excluded instruments)	148.19
Observations	6,053

Note: The table presents results of IV regressions that use ESSA as an instrument for SSA (see equation 1 and 2). The outcome variable is generated using a Box-Cox-transformation $y_i^{transformed} = (y_i^{\lambda} - 1)/\lambda$ with Lambda equal to 4.22 and then standardized to having a mean of zero and a variance of one. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

Finally, as has been discussed in Section 3.3.1 and shown in Figure 3.A.1 in the appendix, the distribution of the receptive vocabulary test score is highly left-skewed. Therefore, Table 3.5 presents results that use transformed test scores as the outcome variable. The transformed receptive vocabulary test scores have a distribution that is much more similar to a normal distribution. The IV estimates remain the same when using this alternative outcome. Similar to the case for the untransformed receptive vocabulary test scores, the effect of SSA is statistically significant, and the effect size is close to a third of a standard deviation (of the transformed test score). As an alternative to using transformed test scores, we also tested whether the results are robust to dropping individuals at the extreme lower end of the untransformed test score distribution. Results are generally not affected by this. For example, if we drop individuals with an untransformed z-score of below -4, the IV estimate is 0.315. If we drop individuals with an untransformed z-score of below -2, the IV estimate is 0.252. Both estimates are statistically significant.

Discussion of the results

Our results show that the effects of school starting age fade away in adulthood for math competencies and for text comprehension but remain significant for receptive vocabulary. This subsection discusses why the long-term impact of SSA differs between competencies. First, we look at the pathways of students in the schooling and vocational systems. This is highly important in the German case since the educational system is characterized by early tracking in school and a strong apprenticeship system. Second, we tie those results together with findings from the literature concerning how individuals acquire competencies. In particular, we discuss how acquiring competencies differs by the particularities of the German educational system.

Table 3.6: IV estimates of school starting age on schooling

	Years of schooling (1)	Highest school degree: low (2)	Highest school degree: middle (3)	Highest school degree: high (4)
IV estimate: SSA	0.5102	-0.1488	0.0283	0.1205
	(0.020)	(0.044)	(0.539)	(0.021)
F-statistic (excluded instruments)	151.0719	151.07	151.07	151.07
Observations	5,920	5,920	5,920	5,920

Note: The table provides IV estimates of the effect of school starting age on the outcomes listed in the first row. The instrument used is ESSA. The estimation sample comprises individuals for whom a valid test score for receptive vocabulary is available. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

In Germany, students are assigned to different school tracks at the age of 10 depending on their abilities and educational performance in primary school. The tracks differ in curriculum, degree of difficulty and abstraction of the material that is covered, in the highest degree attained when leaving school, and in the length of schooling. For the cohorts in our data set, attaining a school degree from the lowest track required 8 or 9 years of schooling (Hauptschule); from the middle track, 10 years (Realschule); and from the highest track, 13 years (Gymnasium). Using data from one out of ten West German states, Puhani and Weber (2007) show that the school starting age has a substantial impact on the assignment of tracks in Germany, meaning that a higher SSA leads to choosing a higher track. Based on the representative NEPS data, we answer this question using all West German states by estimating our main IV model with different educational outcomes as dependent variables, such as the highest school degree

and the highest vocational degree attained. In order to save space, we will not provide results for each of the three samples used in the analysis. Instead, the analysis focuses on the sample for which the receptive vocabulary scores are available, which is not only the sample for which we have found significant results, but it is also the sample with the largest number of observations.

Table 3.7: IV estimates of school starting age on vocational education

	Highest vocational degree: none (1)	Highest vocational degree: apprenticeship (2)	Highest vocational degree: college (3)
IV estimate: SSA	-0.0340	0.0138	0.0202
	0.071	0.787	0.690
F-statistic (excluded instruments)	135.01	135.01	135.01
Observations	5,681	5,681	5,681

Note: The table provides IV estimates of the effect of school starting age on the outcomes listed in the first row. The instrument used is ESSA. The estimation sample comprises individuals for whom a valid test score for receptive vocabulary is available. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

Table 3.6 provides the results of the influence of age at school entry on years of schooling and on the highest school degree attained. Overall, they indicate that being one year older at school entry increases years spent in school by approximately half a year (Column 1). Columns 2 to 4 show that a one-year-higher SSA decreases the probability of attaining the lowest school degree by 15% and increases the probability of attaining the highest school degree by 12%. These findings are very similar to those in Dustmann et al. (2017) on school track at the age of 14.[8] Table 3.7 shows the results for vocational education: distinguishing no vocational degree, having completed an apprenticeship, and obtaining a college degree. SSA only affects the probability of having obtained no

[8] At first glance, our IV estimates appear to be considerably larger than the effects presented in Dustmann et al. (2017), but note that Dustmann et al. (2017) present reduced-form estimates, while we present IV estimates. Our reduced-form estimates are -5.8 ppt for attaining the lowest schooling degree and +4.7 ppt for attaining the highest schooling degree. Table 2 in Dustmann et al. (2017) compares individuals born in different months that differ, on average, in expected school entry age by 0.5 and 0.91 years, respectively. If we rescale the respective estimates to represent one-year differences, the effect sizes from Dustmann et al. (2017) range between -7.8 ppt and -3.8 ppt for attending the lowest school track and between +4.0 ppt and +6.6 ppt for attending the highest school track. Our reduced form estimates fall exactly into these intervals.

vocational degree by 3%, which is much smaller in magnitude than the effect of SSA on the schooling degree. In contrast, SSA has no statistically significant impact on completing an apprenticeship or college.[9] We suggest that the reason for observing SSA effects mostly on school track choice, but not on the highest educational degree, is that SSA effects mirror maturity differences that are biggest, when children are young. When becoming older, maturity becomes less important for explaining educational success and students' real potential becomes visible. The German schooling system is flexible by allowing high ability students from middle school to upgrade their skills after leaving school (Dustmann et al., 2017). These students can even study, if they fulfill some requirements. In addition, high school graduates who have learned of their potential to be lower than expected can abstain from enrolling at university.

Given that we mainly find SSA effects on track choice, we hypothesize that the division into the different school tracks is the potential mechanism for the effects of SSA to persist into adulthood, while selection into different vocational tracks is less important. To shed further light on this hypothesis, we once again estimate the impact of SSA on the receptive vocabulary score but now control additionally for the potential educational mechanisms. Columns 1 and 2 of Table 3.8 control for the highest schooling degree attained and the highest vocational degree, respectively. These findings illustrate that once we control for the schooling degree, the effect of SSA on receptive vocabulary vanishes almost completely. The point estimate drops from 35% of a standard deviation in our main specification to 7% and is no longer statistically significant. In contrast, results in Column 2 show that controlling for the highest vocational degree has a considerably smaller impact. The effect size decreases only modestly; it is still sizeable at 28% of a standard deviation, and it is statistically significant at the 5% level. These results are indicative evidence that SSA affects the track assignment by which it has a long-term impact on receptive vocabulary. However, why does this conclusion not also hold for the other competencies?

Pischke and von Wachter (2008) state that basic math and reading/writing skills are taught in secondary school in Germany, regardless of the track choice. This means that basic skills are taught to all students. The difference between the lower and higher tracks in this regard is mainly that higher-track students learn more advanced and academic

[9] Dustmann et al. (2017) also find that the effects of SSA are much smaller when considering completed education.

Table 3.8: IV estimates of SSA on receptive vocabulary while controlling for mechanisms

	Language: receptive vocabulary (1)	Language: receptive vocabulary (2)
IV estimate: SSA	0.0773	0.2739
	(0.485)	(0.019)
Additional controls for the highest school degree	Yes	No
Additional controls for the highest vocational degree	No	Yes
F-statistic (excluded instruments)	152.10	134.91
Observations	5,920	5,681

Note: The table provides IV estimates of the effect of school starting age on receptive vocabulary. The instrument used is ESSA. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

knowledge. Given the focus of the competency tests on basic skills and their application to everyday problems, advanced knowledge is not captured in the competency tests available in the NEPS data. The basic skills are also necessary for participating successfully and productively in the German labor market (Pischke and von Wachter, 2008). Thus, we consider it plausible that these skills are used and, thereby, practiced regularly after leaving school.

In contrast, receptive vocabulary is learned by exposure to oral or written language.[10] School tracks might differ in terms of vocabulary growth due to students being exposed differently to language activities. Students in the highest track have more opportunities to engage in these activities because they stay in school longer, but also because reading texts or books (including classical literature) is much more frequent in the highest track. The literature has also shown that the number of different and rare words in texts and books matters in terms of increasing individuals' vocabulary growth. The largest variety can be observed in scientific texts (Hayes and Ahrens, 1988). Due to its more academic curriculum, graduates from the highest track are likely to have been exposed to more academic words than lower- or medium-track students.

In a nutshell, school starting age affects the track assignments. The academic track

[10] The scientific literature has not yet reached consensus on whether vocabulary growth occurs mainly incidentally through conversations and reading (Sternberg, 1987) or whether it is transmitted through explanations by teachers or parents or within texts (Biemiller, 2001).

puts much more weight on developing language skills than the middle and the lower school tracks, which might result in a more-refined and larger set of vocabulary used by academic track students. In contrast, there is evidence that academic-track schools are not superior to middle- and lower-track schools when it comes to generating the basic skills gauged in the mathematical literacy and text comprehension tests. The next section is devoted to answering what policy can do to counteract the long-lasting SSA effects on receptive vocabulary.

Absolute vs. relative age effects

As discussed in Cascio and Schanzenbach (2016), the effects documented in most of the literature and in our main results are a mixture of absolute age at school entry and of relative age. Separate identification is important from a policy perspective. For example, take a policy that changes the cut-off date by one month. If only absolute age is important, this policy does not have any impact on those further away from the cut-off, but only on those whose birthday falls between the old and the new cut-off. If those between the old and the new cut-off benefit from the reform, there will also be an improvement when looking at the aggregate of children because no child will be disadvantaged by the reform. In contrast, if relative age is important, this policy also has an impact on children further away from the cut-off because the policy affects the average age of classmates and, thus, the relative age of each student. Furthermore, if only relative age is important and not absolute age, this policy will only influence which child is the oldest and which child is the youngest in the class and, thus, who is hit by the negative effect of being (relatively) young, but it has no effect on the aggregate of children.

Focusing on receptive vocabulary, Table 3.9 provides separate estimates for absolute and relative age effects. Separate estimates are possible because we have variation in cut-off dates between states and over time. This allows us to define separately instruments for age at school entry and for age relative to the child that should be the youngest within a cohort, given all children follow the regulations. However, both instruments are highly correlated, with a correlation coefficient of 0.91. This correlation makes separate identification somewhat problematic, and we emphasize that the separate effects should not be over-interpreted. According to our findings, only absolute age effects are relevant for receptive vocabulary. In contrast, the point estimate of relative age has a negative

sign, is much smaller and is far from being statistically significant. This finding is similar to that of Cascio and Schanzenbach (2016), who also report that absolute age is more important than relative age.

Table 3.9: IV estimates of absolute and relative school starting age

	Language: receptive vocabulary
IV estimate: absolute SSA	0.4148
	(0.039)
IV estimate: relative SSA	-0.0933
	(0.372)
F-statistic (excluded instruments)	72.13
Observations	6,053

Note: The table provides IV estimates of the effects of absolute and of relative school starting age on receptive vocabulary. Estimation accounts for clustering at the state level. Small-sample-adjusted p-values are shown in parentheses.

3.6 Conclusion

The previous literature has shown that scores of competency tests administered to school children are influenced by the age of children at school entry. While there is evidence that these effects become smaller as the children grow older, little is known about whether the effects fade away completely or remain important long after leaving school. We find no evidence that the effects of school starting age on math competencies and text comprehension are still relevant in adulthood, although they are considerable when children are in school. These results are also in line with the previous literature, which has shown no or only small long-term effects of SSA on wages and employment (e.g., Fredriksson and Öckert, 2014; Larsen and Solli, 2017; Dustmann et al., 2017). Assuming that basic competencies in math and text comprehension matter for labor market success, the absence of long-run SSA effects on these competencies could explain the absence of long-run effects of SSA on labor market success. Hence, we conclude that our results do not provide reasons for policy actions.

In contrast, for receptive vocabulary, the effect of school starting age remains large and statistically significant in the longer run. Our findings suggest that the long-run effect on receptive vocabulary is due to Germany's tracking system, which sorts children

into different school tracks at an early age. Given that, this effect is due to absolute rather than to relative age, recent policies that shift the cut-off to an earlier date, making some children older by one year at school entry, should lead to an improvement in the average receptive vocabulary competencies even in adulthood. Of course, the benefits of such reforms need to be contrasted with their social costs (e.g., the cost of one more year in childcare) and private costs (e.g., of entering the labor market one year later).

3.A Appendix

Table 3.A.1: Studies considered in Figure 3.1

	Grade level										
	0	1	2	3	4	5	6	7	8	9	10
	Math test scores										
Attar and Cohen-Zada (2018)						x			x		
Bedard and Dhuey (2006)				x	11x				18x		
Black et al. (2011)											x
Cook and Kang (2018)				x							
Datar (2006)	x										
Elder and Lubotsky (2009)	2x	x		x		x			x		
Koppensteiner (2018)						x					
Lubotsky and Kaestner (2016)	2x										
McEwan and Shapiro (2008)					x				x		
Nam (2014)							x	x	4x		
Peña (2017)				x	x	x	x	x	x	x	
Ponzo and Scoppa (2014)					x				x		
Smith (2009)					x			x			x
	Reading/writing test scores										
Attar and Cohen-Zada (2018)						x			x		
Cook and Kang (2018)				x							
Datar (2006)	x										
Elder and Lubotsky (2009)	2x	x		x		x			x		
Lubotsky and Kaestner (2016)	3x										
McEwan and Shapiro (2008)					x						
Nam (2014)							x	x	x		
Peña (2017)				x	x	x	x	x	x	x	
Ponzo and Scoppa (2014)					x						
Puhani and Weber (2007)					x						
Smith (2009)					2x			2x			2x

Notes: Those studies that include several estimates per grade level report either separate estimates for different countries, such as in Bedard and Dhuey (2006), or different test scores, such as separate tests for reading and writing.

Table 3.A.2: Month of school start by state

State	Month of school start
BW	1950-1951: September, 1952-1966: April, 1966-1994: August
BY	1950-1994: August
HB	1950-1966: April, 1967-1994: August
HH	1950-1966: April, 1967-1994: August
HE	1950-1966: April, 1966-1994: August
NI	1950-1966: April, 1967-1994: August
NW	1950-1966: April, 1966-1994: August
RP	1950-1966: April, 1966-1994: August
SL	1950-1958: August, 1959-1966: April, 1966-1994: August
SH	1950-1966: April, 1967-1994: August
BE	1950-1951: August, 1952-1966: April, 1967-1994: August

Source: State-specific laws and legislation determining the month of school start.
Note: In some states, the school year started two times during 1966 (short school year).

Table 3.A.3: Cut-off date by state

State	Cut-off date
BW	1950: 31.12., 1951: 31.5., 1952: 31.3., 1953-1963: 15.4., 1964-1966: 31.12., 1966-1994: 30.6.
BY	1950-1968: 30.9., 1969-1994: 30.6.
HB	1950-1965: 31.3., 1966: 31.5. & 30.11., 1967: 1.7., 1968-1994: 30.6.
HH	1950-1961: 31.3., 1962-1966: 31.12., 1967-1994: 30.6.
HE	1950-1956: 30.6., 1957-1961: 31.3., 1962-1965: 31.12., 1966: 31.3. & 30.11., 1967-1994: 30.6.
NI	1950-1955: 30.6., 1956-1966: 31.3., 1967-1994: 30.6.
NW	1950-1960: 30.6., 1961-1965: 31.3., 1966: 31.3. & 30.11., 1967-1994: 30.6.
RP	1950-1952: 30.6., 1953-1965: 31.3., 1966: 31.3. & 30.11., 1967-1994: 30.6.
SL	1950-1954: 31.12., 1955-1957: 30.9., 1958: 31.12., 1959-1956: 31.3., 1966: 31.3. & 31.12., 1967: 30.9., 1968-1994: 30.6.
SH	1950-1955: 30.6., 1956-1963: 31.3., 1964-1965: 31.12., 1966: 31.12. & 30.11., 1967-1994: 30.6.
BE	1950-1951: 31.12., 1952-1955: 30.6., 1956-1966: 31.3., 1967-1994: 30.6.

Source: State-specific laws and legislation determining the cut-off date of school entry for children.
Note: In some states, the school year started two times during 1966, resulting in two cut-off dates.

Figure 3.A.1: Distribution of the standardized test scores

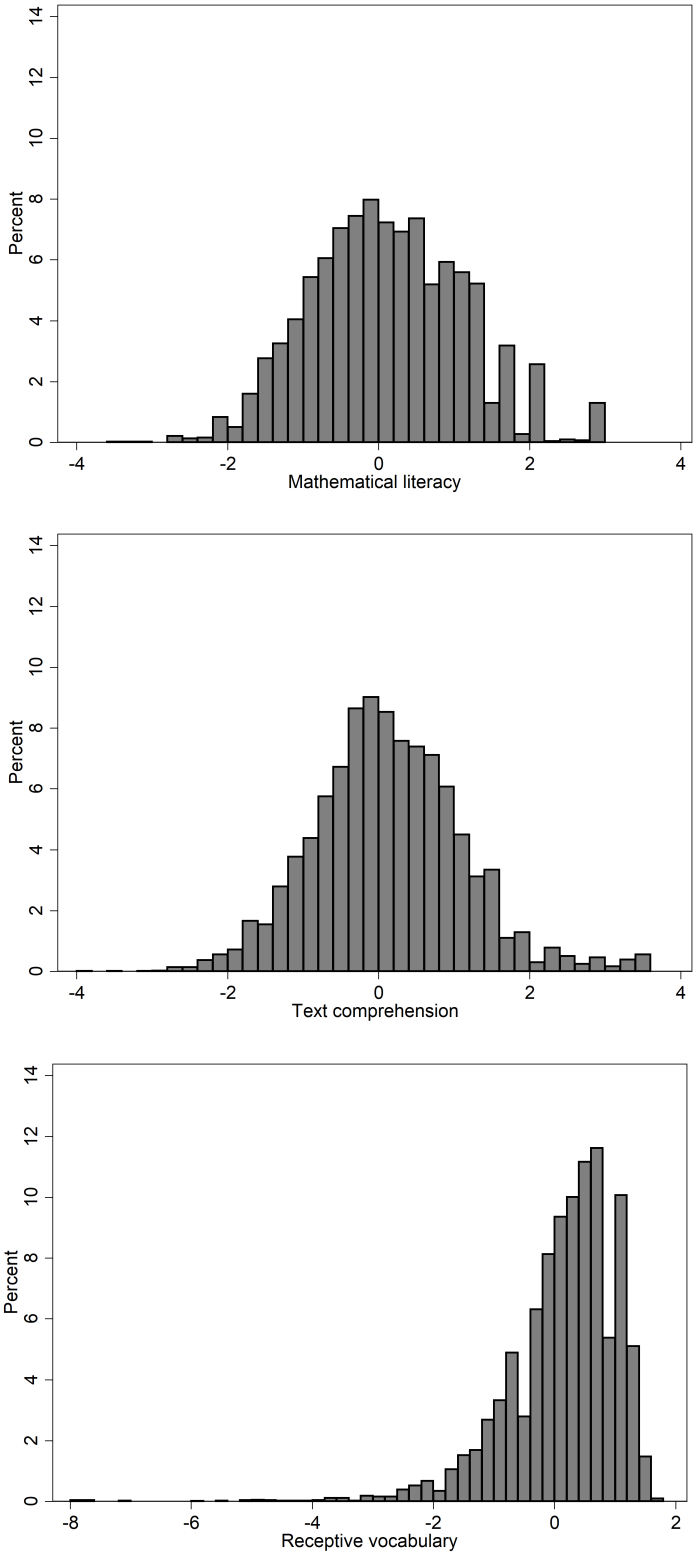
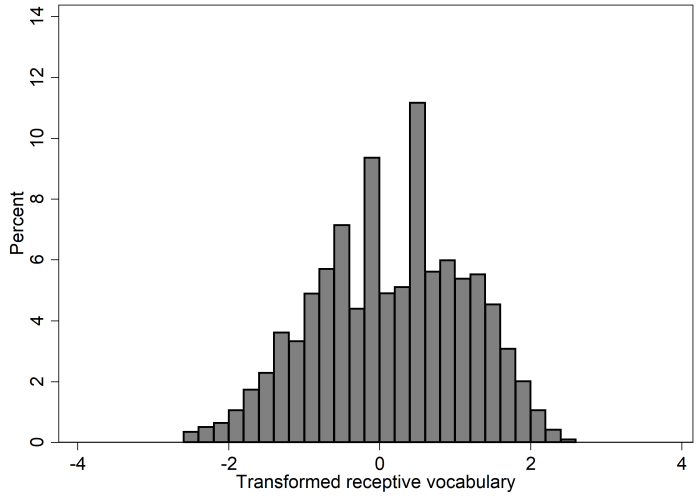
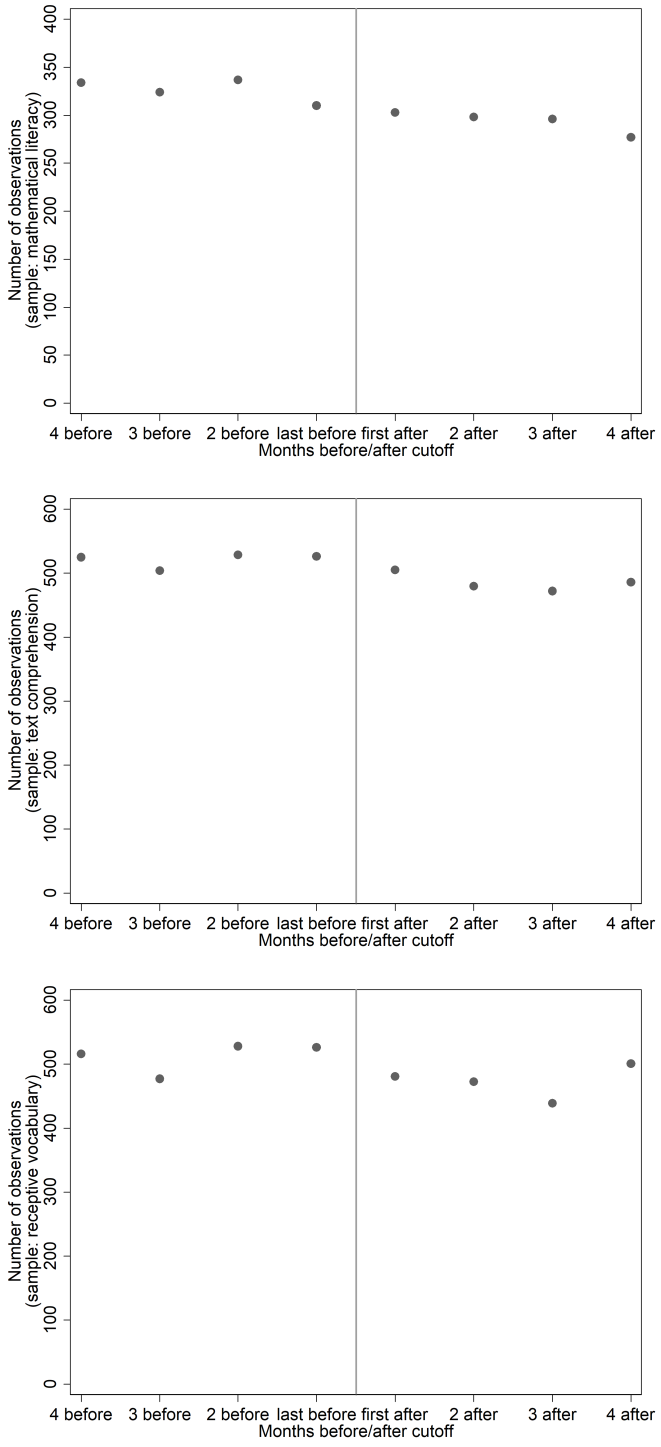


Figure 3.A.2: Distribution of the transformed receptive vocabulary test scores



Note: The transformed test scores are generated using a Box-Cox-transformation $y_i^{transformed} = (y_i^{\lambda} - 1)/\lambda$ with Lambda equal to 4.22. The transformed test scores are then normalized to having a mean of zero and a variance of one.

Figure 3.A.3: Number of observations by distance to cut-off



Note: Figures are for samples with information on mathematical literacy (top), text comprehension (center), and receptive vocabulary (bottom).

Chapter 4

Now You See Me! Ethnic Discrimination in the Market for Apprenticeships*

4.1 Introduction

It is well known that ethnic discrimination exists in all areas of life, extending from social markets (Tjaden et al., 2018) and the housing market (Ewens et al., 2014) to the labor market (Bertrand and Mullainathan, 2004). The understanding of discrimination is important to promote economic and social integration and equal opportunities for every individual in a society. Its importance becomes even more clear when realizing that discrimination directly affects the prospects of an individual in the labor market, not only directly via wages and employment opportunities but also through educational performance, e.g. in the form of self-fulfilling prophecies (Glover et al., 2017).

Observed discrimination in the labor market hiring process is often seen as the result of limited information about applicants. This leads to hiring decisions based on perceived differences in the average productivity of different groups. When this is the case, additional information such as references, training program certifications and work experience are able to decrease discrimination. However, in case of young individuals in the transition from school to the first job, there exist only limited possibilities to

*This chapter is based on joint work with Patrick Nüß (Macroeconomic Policy Institute (IMK), Christian Albrechts-University Kiel).

provide the required information. It is therefore unclear whether youths face similar levels of discrimination when applying for apprenticeships as workers do who apply for regular jobs.

Survey data reveal that despite the fact that young people with migration background send more applications, they are less likely to be invited to a job interview (49%) than their German counterparts (62%) (Beicht, 2016). Furthermore, applicants without German citizenship have only half the apprenticeship entry rate compared to those with German citizenship (BMBF, 2018). The BMBF (2018) explains this with differences in occupational preferences, regional labor markets, limited proficiency in the German language and fewer internships. But what is the role of discrimination? Given the importance of the school-to-work transition for future labor market prospects, it is important to understand the relevance of discrimination and to search for solutions in the school-to-work transition.

To solve the issue of discrimination in the first stage of the hiring process, anonymous application procedures are a common and successful solution (Goldin and Rouse, 2000; Krause et al., 2012). However, this could simply delay discrimination to a later stage. Furthermore, ethnic discrimination is more likely to be observed in small firms (Kaas and Manger, 2012) where anonymous application procedures are difficult to implement. Giving this prevailing circumstance in the market for apprenticeships, there is a need for alternative solutions to anonymous procedures which are easily accessible enough to solve discrimination in the school-to-work transition. One such solution which is gaining popularity involves video applications. These allow a deeper insight into the personality and motivation of the applicant.

In this paper a correspondence experiment is used to study ethnic discrimination in the hiring market for apprenticeships in Germany. We send fictitious applications with German and Arabic-sounding names to real vacancies, augmenting applications with a short motivational video. These video introductions of applicants serve as a simple way of extending employer's insight into the personality and motivation of an applicant. This is done with the intention of decreasing potential statistical discrimination and to test the limitations due to attention discrimination (Bartoš et al., 2016).

We find that applications with German-sounding names are on average twice as likely to receive an invitation for a job interview. Videos of applications with German-sounding names are about 50% more likely to be viewed than those from applicants

with Arabic-sounding names. However, callbacks for applications with a video increase by 33% for Arabic-sounding names compared to applications with Arabic-sounding names without video. A similar effect for videos can not be found for German-sounding names. Furthermore, despite the fact that videos only have observable effects on Arabic-sounding names, the videos of German-sounding names are viewed 37% more often compared to the Arabic-sounding names, which is in support of attention discrimination.

The remainder of the chapter is organized as follows. Section 4.2 gives an overview of relevant literature. Section 4.3 describes the experimental design and callback measures. Section 4.4 presents the results and provides robustness and heterogeneity analysis, Section 4.5 concludes.

4.2 Literature Review

For the labor market there exists comprehensive research on the existence of discrimination for a range of countries.[1] Due to limitations of administrative and survey data, a common approach to isolate discrimination from alternative explanations is the use of correspondence experiments. One prominent experiment was conducted by Bertrand and Mullainathan (2004) who send fictitious applications with African-American- compared to white-sounding names to job vacancies. They provide evidence for lower callbacks of African-American-sounding names compared to their white-sounding counterparts and therefore for ethnic discrimination in the United States.

The results that native-sounding names receive significantly more callbacks has in the following years also been found for European countries when comparing Turkish-, Arab- or Middle Eastern-sounding to native-sounding names (Weichselbaumer, 2016b; Baert et al., 2015; Kaas and Manger, 2012; Carlsson and Rooth, 2012; Rooth, 2010), including evidence for ethnic discrimination in Austria (Weichselbaumer, 2016b) and Germany (Kaas and Manger, 2012; Goldberg et al., 1996).

According to the literature, ethnic discrimination can be explained by the existence of taste-based discrimination (Becker, 1957) and statistical discrimination (Arrow, 1973). Taste-based discrimination argues for discrimination based on the employer's

[1] Excellent overviews of correspondence experiments on hiring discrimination can be found in Baert (2018), Bertrand and Duflo (2017), Zschirnt and Ruedin (2016), Rich (2014) and Riach and Rich (2002).

prejudice or distaste for the cultural background of a group membership. As an alternative explanation, statistical discrimination is the result of asymmetric information. Because of limited information about the applicants, employers might use group memberships as a signal for perceived differences of the average productivity related to the group.

If statistical discrimination instead of taste-based discrimination is the driving force, extending the access to information about applicants should decrease or even solve discrimination. Previous findings for several markets indicate the dominance of statistical discrimination as a driver for ethnic discrimination. Altonji and Pierret (2001) show that if more information is available to the employer, the applicants' wage is more dependent on their productivity. By analysing the rental market Ewens et al. (2014) find discrimination related to the neighbourhood's racial composition consistent with statistical discrimination. Further evidence for statistical discrimination can also be found for the sports card market (List, 2004) or online market places (Doleac and Stein, 2013; Nunley et al., 2011), as well as for carpooling marketplaces (Tjaden et al., 2018). In addition, in the case of a virtual labor market for football players, Bryson and Chevalier (2015) provide evidence against taste-based discrimination.

While most of the literature seems to confirm the existence of statistical discrimination, recent results show that statistical discrimination can not explain ethnic discrimination alone. For example, Oreopoulos (2011) finds that native-sounding names receive significantly more callbacks than foreign-sounding names. He also discovers that discrimination does not decrease when adding additional information to the applications which would be evidence of statistical discrimination. Employers view language barriers as the main reason for discrimination, yet when providing information about a high level of language proficiency, discrimination remains. Further investigation provides evidence for hiring discrimination in case of religious applicants (Valfort, 2017) or indirectly through appearance, for instance should the applicant be wearing a headscarf (Weichselbaumer, 2016a).

Beside the debate of statistical and taste-based discrimination, Bartoš et al. (2016) find evidence for attention discrimination as an alternative explanation for discrimination. They show that employers choose different levels of attention for applicants of an ethnic minority compared to those of an ethnic majority in different markets, including the labor market. This is supported by a lab-in-the-field experiment using eye-tracking

technology by Lahey and Oxley (2018) who find that screeners spend less time viewing a resume if it is of a young and black applicant, which implies that even the existence of additional information does not guarantee its usage.

Focusing on experiments in German speaking labor markets, there is evidence for statistical discrimination (Kaas and Manger, 2012), taste-basted discrimination (Weichselbaumer, 2016a) as well as attention discrimination (Bartoš et al., 2016). While Weichselbaumer (2016a) and Bartoš et al. (2016) focus on adults with existing work experience, the experiment of Kaas and Manger (2012) focuses on young university students, applying for an internship. In another experiment in the market for apprenticeships in Germany, Kübler et al. (2019) provide evidence of increasing callbacks for young people who have gained additional vocational training experience before beginning an apprenticeship. Our study focuses on ethnic discrimination in school to work transition in the market for apprenticeships, with lower educational achievements than Kaas and Manger (2012) and without the signal of university attendance.

Previous labor market research on ethnic discrimination focuses on statistical discrimination related to qualifications based on grades (Veit and Yemane, 2018), language proficiency (Oreopoulos, 2011) and recommendation letters (Kaas and Manger, 2012). Our research benefits from research on school to work transitions in Germany that considers the relevance of non-cognitive skills as a potential solution for ethnic discrimination. Solga and Kohlrausch (2013) study low-achieving German youths transitioning into apprenticeships. They observe increasing job prospects if employers can observe the motivation and behavior of these applicants. This suggests non-cognitive skills are more important than grades for the low-skilled youths who apply for apprenticeships. Protsch and Solga (2015) conduct a field experiment and provide evidence for the relevance of both cognitive and non-cognitive skills, with a higher relevance of non-cognitive skills.

Concluding, we suspect to observe an overall high level of ethnic discrimination in the market for apprenticeships due to the fact that individuals have limited work experience and qualifications. Due to the higher relevance of non-cognitive skills in this market, augmenting applications by a short motivational video might be a simple and suitable way of solving ethnic discrimination by extending the employers' potential insights into the personality and the motivation of an applicant. The success is a priori unclear

since such videos might also be subject to the potential risk of attention discrimination (Bartoš et al., 2016).

4.3 The Experiment

To analyze ethnic discrimination in the market for apprenticeships in Germany we conduct a correspondence experiment to analyze employer attitudes in a real world context. Field experiments have been applied to a wide range of settings ranging from ethnic discrimination (Bertrand and Mullainathan, 2004) to gender discrimination (Riach and Judith, 2006) and in recent years also to employers' attitudes toward tattoos (Jibuti, 2018) and unemployment (Nüß, 2018).

The basic idea is to send one or more fictitious job applications, similar in all aspects except one characteristic, to real vacancies and then monitor firms' callbacks. This concept is best suited to analyze employers' attitudes towards applicants based on these characteristics. This is possible since the relevant characteristic can be isolated from the applicant's motivation and skill. The latter are not commonly observable by the researcher but observable to the employer during the hiring process. A correspondence therefore allows full access to all information available during the hiring process. With all productivity relevant characteristics equal and the unobservables isolated, monitoring firms' callbacks allows the causal effect of a randomly assigned characteristic in the hiring process to be identified. This, in the end, rules out alternative explanations for different labor market prospects mentioned in BMBF (2018).

4.3.1 Experimental Design

The experiment was carried out between February and May 2018 in 26 regions in Germany. The timing results from the fact that application activity increases in the mid February, directly after schools hand over the midterm certificates and the beginning of the apprenticeship in August/September each year. Three apprenticeship programs are chosen, these are automotive technician, sales and office clerk. The occupational choices result from the top 5 most popular male and female apprenticeships and by their overall availability within Germany (BIBB, 2018). There are four identities, two each with German- and Arabic-sounding names.

Vacancies are collected from three large websites providing vacancies in Germany, two of which specialize in vacancies for apprenticeships. This increases the external validity of the experiment and opens the door for an analysis of regional specific aspects.

Three applications are sent via email to each firm. These applications consist of a cover letter, a resume and the school certificates with above average grades (1.9) and comparable grades by subjects.[2] The name and figure is randomly assigned. All applicants are 16 years old and are currently completing the 10th grade (secondary school). Their resumes include work experience in the relevant occupation resulting from a completed internship. Additionally, one of the three applications is augmented by a short 16-19 second video introducing the applicant. The video is then accessible by a link to a private YouTube channel (see Figure 4.A.1 for a screenshot of a video).

When a video is assigned to an application, the link is included in both the email and the cover letter. First, the link is included in the initial email introducing the candidate and describing the attachments. This is in order to allow for immediate attention for the application. Second, the link is also included in the cover letter. By doing so we ensure that the video may be considered in a later stage of the hiring process when the applications are screened in more detail and the initial email might already be out of mind. In both cases the applicant mentions that he added a video in order to allow for a first personal impression.

4.3.2 Measurement of Callbacks

To measure the callbacks, each of the resulting four identities is linked to an individual's phone number and an email address. This allows for the receipt of callbacks by email and by telephone via a linked voice mail. To minimize any inconvenience caused to firms, emails and voice mails are checked on a daily basis and every callback is immediately declined. A positive callback in the literature follows typically two categorizations, both considered in the following analysis. A positive callback following the strict definition (Callback Category 1) includes all callbacks explicitly inviting the applicant to a job interview for the apprenticeship. A positive callback based on callbacks in a broader sense (Callback Category 2) includes every kind of request for additional information, alternative apprenticeship offers, invitations for an internship and requests to contact the employer.

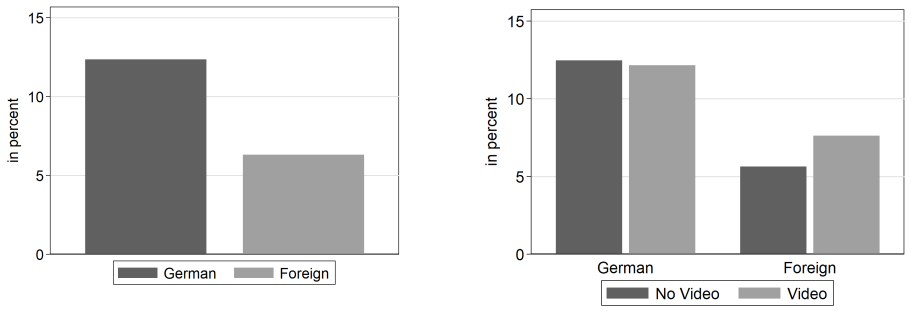
[2] Grades in Germany vary from 1 (excellent) to 6 (insufficient).

4.4 Results

4.4.1 Descriptive results

Table 4.1 gives an overview of the final dataset. The sample consists of 3828 applications sent to 1276 firms. The most frequent observation are firms with 6 to 50 employees (55%) in West Germany (71%), with a request for 10 years of educational attainment (64%) for applicants. The data set consists of 48% apprenticeships for office clerks and an equal share of 26% for sales and automotive technicians. A majority of 60% of the apprenticeships are advertised in February and March. Furthermore, we observe an approximately equal share of male and female human resources managers responsible for the hiring process.

Figure 4.1: Callback rate



(a) Overall **(b)** Videos

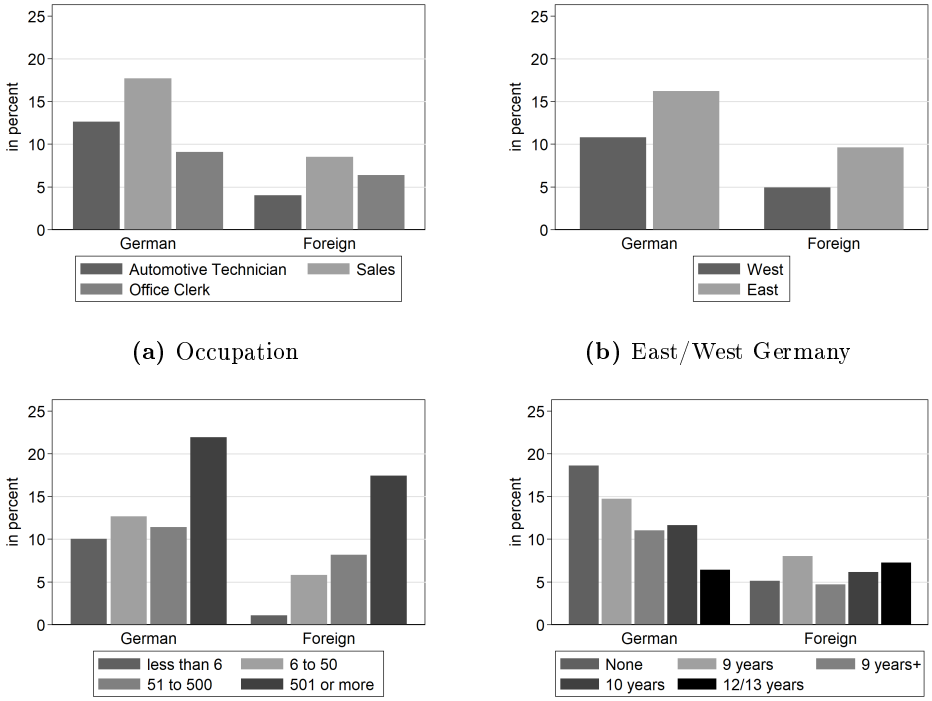
Note: Callback category 1.

Overall 9% of all applicants are invited to a job interview (Callback Category 1) and 16% get at least a positive callback in a broader sense (Callback Category 2). This covers 17% of firms inviting at least one applicant to a job interview (28% for Callback Category 2). In the following detailed analysis the focus is on an interpretation based on Callback Category 1. However, the results are overall consistent in size and significance with callback category 2 and available in the appendix (Table 4.A.1 and Table 4.A.2). When comparing the 9% overall invitations to job interviews of Table 4.1 with Table 4.A.1 and Figure 4.1a we see that invitations are unequally distributed between German- and Arabic-sounding names. On average applicants with a German-sounding name are

Table 4.1: Summary

Variable	Mean	Std. Dev.	Min.	Max.	N
Callback of category 1 (C1)	0.093	0.291	0	1	3828
Callback of category 2 (C2)	0.163	0.370	0	1	3828
Firms with at least one callback (C1)	0.165	0.371	0	1	1276
Firms with at least one callback (C2)	0.278	0.448	0	1	1276
Video	0.333	0.471	0	1	3828
Foreign	0.501	0.500	0	1	3828
East Germany	0.287	0.452	0	1	3828
Job begin in August	0.715	0.452	0	1	3828
Job begin in September	0.285	0.452	0	1	3828
No. of open positions	1.196	0.584	1	5	3828
HR female	0.495	0.500	0	1	3828
HR gender unknown	0.045	0.207	0	1	3828
Occupation:					
Automotive Technician	0.263	0.440	0	1	3828
Office Clerk	0.473	0.499	0	1	3828
Sales	0.264	0.441	0	1	3828
Firm size (employees):					
Less than 6	0.144	0.351	0	1	3828
6 to 50	0.543	0.498	0	1	3828
51 to 500	0.269	0.443	0	1	3828
501 or more	0.044	0.205	0	1	3828
Educational Demand (years of schooling (German degree)):					
None (*Keine Anforderung*)	0.057	0.232	0	1	3828
9 years (*Hauptschulabschluss*)	0.176	0.381	0	1	3828
9 years+ (*Erw. Hauptschulabschluss*)	0.100	0.299	0	1	3828
10 years (*Mittlere Reife*)	0.636	0.481	0	1	3828
12/13 years (*Hochschulreife*)	0.031	0.172	0	1	3828
Month of job posting:					
February	0.369	0.483	0	1	3828
March	0.227	0.419	0	1	3828
April	0.216	0.411	0	1	3828
May	0.188	0.391	0	1	3828

Figure 4.2: Callback rate



(a) Occupation

(b) East/West Germany

(c) Firm size

(d) Educational Demand

Note: Callback category 1.

invited to a job interview in 12.4%, compared to 6.3% for Arabic-sounding names. This implies that applicants with German-sounding names are twice as likely to be invited to a job interview compared to their equally qualified Arabic counterparts, significant on the 0.1% significance level. This observation is in line with the BMBF (2018) observation of twice as high apprenticeship entry rates of Germans compared to non-Germans.

Looking at applications with and without videos reveals that invitations to job interviews do not change for the German group (Figure 4.1b). In the foreign group, however, the invitations to job interviews increase by 2 percentage points compared to applications without video. This corresponds to a 36 percent increase in callbacks, yet insignificant.

Figure 4.2 provides further insight into differences in invitations to job interviews and discrimination, in consideration of (a) occupational variation, (b) regional differ-

ences, (c) firm size and (d) firms' educational requirements. As before, statistical test regarding this variation is available in the appendix (Table 4.A.1 and Table 4.A.2). From Figure 4.2a we observe that of the three occupations, automotive technician applicants with Arabic-sounding names are the most disadvantaged since applicants with German-sounding names receive more than 3 times more invitations to job interviews. For sales positions, Germans are twice as likely to be invited to a job interview. For office clerks, we observe the lowest discrimination with German-sounding names being 1.5 times as likely to be invited to a job interview compared to their Arabic-sounding counterparts.

In general, regional variation plays an important role for discrimination. According to Carlsson and Rooth (2012) regions where employers have a more negative attitude towards minorities have lower callbacks for minority groups. This has implications for our study since differences in attitudes towards ethnic minorities may also vary between former East and West German states. Figure 4.2b therefore considers differences in discrimination related to West and East Germany. Interestingly we observe on average 5 percentage points higher callbacks in East Germany for German- as well as Arabic-sounding names. As a result, in West Germany, German-sounding names are 2.2 times more likely to be invited to a job interview than Arabic-sounding names, compared to 1.7 times in East Germany. Contrary to what could have been expected, this implies lower discrimination in East Germany compared to West Germany. The most likely explanation for this is that the skills shortage in East Germany is more pronounced than in West Germany and firms are therefore keen to avoid the risk of a vacant apprenticeship.[3]

It is known from Kaas and Manger (2012) that ethnic discrimination is likely to be higher for small firms. This is consistent with Figure 4.2c that provides evidence for decreasing discrimination with increasing firm size. Most strikingly, there are almost no invitations to job interviews for Arabic-sounding names observable in firms with less than 6 employees.

The observed pattern for educational requirements of firms in Figure 4.2d is also very interesting. Intuitively one would expect ethnic discrimination to be lowest when employers have no educational requirements for applicants. However, employers with

[3]In 2018 the applicants to positions ratio (averaged over the three occupations and regions) is 1.5 in West but only 1.3 in East Germany.

no requirements about an applicant's education discriminate most and are 3 times more likely to invite German-sounding names to a job interview. With increasing requirements, discrimination also decreases. While the applicants have 10 years of schooling and are not able to fulfill the requirements of some firms, firms still invite 7% German- as well as Arabic-sounding names to job interviews with no detectable discrimination. Regarding Figure 4.2d another interesting pattern is observed. Callbacks for German-sounding names decrease with increasing requirements of firms, leading to the observed decrease in discrimination. On the other hand, invitations for Arabic-sounding names seem to be independent of firms requirements and are indistinguishable from firms where the applicants do not fulfill the demanded educational requirements.

4.4.2 Multivariate analysis

To ensure the robustness of our findings and to control for all potential variables, the analysis is extended by a linear probability model, again focusing on invitations to job interviews (callback category 1). As before, robustness checks for callback category 2 are available (Table 4.A.3) supporting the findings of the following analysis.

In Table 4.2 for callback category 1 the baseline estimation in column (1) includes a dummy variable for foreign ethnicity and another dummy variable for the effect of the attached video. Very similar to the previous results, here applicants with a Arabic-sounding name have a 6.1 percentage point lower probability of an invitation to a job interview compared to applicants with German-sounding names, significant on the 0.1% level. With an effect size below 1 percentage point, however, we find no evidence for a relevant or significant effect of the attached video. In a step-by-step extension in column (2)-(7), we further control for (2) application characteristics, (3) picture and name, (4) firm size, (5) city, (6) occupation as well as (7) additional firm information. We further check robustness by (8 and 10) firm fixed effects and specifications (9 and 10) allowing for effect heterogeneity for the video based on ethnicity.

By extending the set of control variables, we observe a robust and an increased level of discrimination against Arabic-sounding names of 7.5 to 8.5 percentage points, with all specifications significant at the 0.1% significance level. A glance at the effect of the motivational video provides no change in the effect or significance for column (2)-(8). However, when allowing for effect heterogeneity for the video based on ethnicity, even the percentage point effect size of the video converges to zero for German-sounding

Table 4.2: Probability of a callback of category C1 (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Foreign	-0.0605***	-0.0608***	-0.0754***	-0.0759***	-0.0757***	-0.0754***	-0.0755***	-0.0735***	-0.0841***	-0.0788***
	(0.0094)	(0.0094)	(0.0171)	(0.0171)	(0.0170)	(0.0169)	(0.0169)	(0.0159)	(0.0176)	(0.0162)
Video	0.0084	0.0084	0.0095	0.0095	0.0095	0.0095	0.0095	0.0092	-0.0033	-0.0014
	(0.0101)	(0.0101)	(0.0101)	(0.0100)	(0.0099)	(0.0099)	(0.0099)	(0.0086)	(0.0157)	(0.0148)
Video × Foreign									0.0255	0.0211
									(0.0199)	(0.0195)
Layout, Order		✓	✓	✓	✓	✓	✓	✓	✓	✓
Picture, Name			✓	✓	✓	✓	✓	✓	✓	✓
Firm size				✓	✓	✓	✓		✓	
City					✓	✓	✓		✓	
Occupation						✓	✓		✓	
Educational Demand							✓		✓	
HR Gender: female, unknown							✓		✓	
Firm FE								✓		✓
N	3828	3828	3828	3828	3828	3828	3828	3828	3828	3828
adj. R^2	0.010	0.011	0.013	0.020	0.041	0.048	0.051	0.491	0.052	0.491

Notes: OLS. Callback category 1. Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, robust standard errors (in parentheses) for (1)-(7) and (9), cluster robust standard errors adjusted for clusters at the firm level for (8) and (10).

names. As expected based on Figure 4.1a, a sizable coefficient for the interaction term of video and foreign ethnicity is found. We observe an increase in the video effect for Arabic-sounding names to between 2 and 2.5 percentage points, however, remaining insignificant.[4] The three depicted coefficients in (10) are jointly significant.[5] This means that a video increases the probability of a callback for ethnic minorities by 38 percent. We interpret the results as weak evidence for the hypothesis of videos reducing statistical discrimination by providing additional information about candidates.

Results for callback category 2 in Table 4.A.3 reveal a much smaller interaction coefficient for video and ethnicity which is 1.3 percentage points (column (10)). This suggests that the video is more relevant for actual invitations of ethnic minorities to job interviews rather than affecting employers asking for additional information or calling back with other offers such as internships.

As was stated in the literature review, discrimination can also be the result of attention discrimination based on ethnicity. Statistical discrimination would predict that additional information accessible to the employer is used to solve asymmetric information and to avoid decision making based on prejudices about a group. Contrary to this attention discrimination implies that employers spend less time screening resumes of minority groups, thereby ignoring easily accessible information. In order to investigate the potential relevance of attention discrimination for our videos, Table 4.3 provides insight into employers screening behavior with regard to the attached video links.[6]

In total 1276 applications with a link included are sent. The link is included in the initial email as well as the cover letter. Employers therefore are confronted with the video in the initial stage and during a later, more detailed screening of the applications. The videos have an average duration of 17 seconds, however the average watch time of the video is 20% longer than the video which holds for German- and Arabic-sounding names.[7] This suggests that employers seriously consider the video while watching it and in some cases even watch parts of it again.

[4] This reduces the remaining ethnic discrimination to $-7.88 + 2.11 = 5.77$ percentage points.

[5] The F-statistic, which reports the results from a joint significance test with the null hypothesis $H_0 : Foreign = 0$, $Video = 0$ and $Video \times Foreign = 0$, is 8.5 and reports a p-value of 0.0000 thus rejecting H_0.

[6] However, due to data protection legislation video links are not personalized and can therefore not be linked to viewing behavior of individual employers.

[7] With a 1.2 second difference between German and Arabic individuals.

Table 4.3: Video Statistics

	German			Foreign			Difference	
	Sum	Mean	SD	Sum	Mean	SD	b	t
Views	702			478				
Video links sent	634			642				
Views / Video links sent	1.11			0.74				
Watch time (seconds)		19.94	4.59		21.13	5.19	-1.19***	(-4.04)
Video duration (seconds)		16.25			17.50			
Watch time / Video duration		1.23	0.28		1.21	0.30	0.02	(1.15)
Observations		702			478		1180	

Notes: Mean individual watchtime calculated from daily watchtime and views reported by youtube.com; Total watchtime (in minutes): 233.32 (German) and 168.33 (Foreign); Period: 2018/3/25-2018/7/1; Differences of means and t-statistics reported; $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Out of 634 applications including a video for German-sounding names, the video was viewed 702 times implying the video of German-sounding names was on average viewed 1.1 times. Contrary to this, out of 642 applications that included a video for Arabic-sounding names, the video was viewed 478 times implying on average that the video was viewed 3/4 of the time. Therefore, despite the fact that videos only had observable effects on Arabic-sounding names, the videos of German-sounding names were viewed 37% more often compared to the Arabic-sounding names, which is in support of attention discrimination (Lahey and Oxley, 2018; Bartoš et al., 2016). This results in a dilemma that additional information as well as effort is not equally valued for all applicants as already observed in the unequal treatment related to the firms' educational requirements (Figure 4.2d).

4.4.3 Heterogeneity and Robustness

Complementary to the analysis of ethnic discrimination and the video solution to diminish discrimination, we complement the analysis with additional robustness checks and a heterogeneity analysis that has attracted attention in Figure 4.2. In the following the regional relevance and market tightness is addressed first, secondly, characteristics

and requirements of firms and for further robustness, the time until the begin of the apprenticeship and school background.

Labor market tightness

The most surprising observation in Figure 4.2 is the observation of a higher number of invitations to job interviews and lower discrimination in East compared to West Germany. This is surprising for two reasons. First, Carlsson and Rooth (2012) provide evidence for an association of a regionally more negative attitude to minorities and observed hiring discrimination, which is in support of taste-based discrimination. The rising popularity of the right-wing political party AfD (*Alternative für Deutschland*) in the 2017 election[8] especially in former Eastern German states is expected to reflect such a negative regional attitude towards minorities which allows us to test for taste-based discrimination.

The same is expected regarding statistical discrimination. As shown in the literature on ethnic discrimination, discrimination decreases in strong labor markets with low unemployment where filling a vacancy becomes more difficult (Baert et al., 2015; Carlsson et al., 2018). As a consequence, one would expect higher discrimination in East Germany, which is characterized by higher unemployment than West Germany. However, this does not seem to be the case for the apprenticeship market.

In the following we therefore consider the regional labor market tightness for apprenticeships for all tested occupations and regions. The regional applicants to positions ratio, differs substantially for the 26 city regions (which correspond to 17 labor market regions) and range from 0.58 to 2.27 for the three occupations.[9] Therefore, effects for East/West Germany should disappear when controlling for labor market tightness.

Table 4.4 considers the regional access for applicants to the tested occupations. Again, when comparing East and West Germany, we observe 5 percentage points more invitations to job interviews in East Germany. Contrary to what the success of the right-wing populist party AfD might suggest, we do not observe higher discrimination in East Germany. As expected, when considering the labor market tightness for apprenticeships, we observe decreasing invitations to job interviews when a higher number of

[8] `https://www.bundeswahlleiter.de/en/bundestagswahlen/2017/ergebnisse.html`

[9] See Bewerber-Ausbildungsstellen-Relation (Oktober 2017 bis September 2018) `https://arbeitsmarktmonitor.arbeitsagentur.de/`

Table 4.4: Probability of a callback of category C1: East Germany, labor market tightness and AfD votes (OLS)

	(1)	(2)	(3)	(4)
Foreign	-0.0821***	-0.1083**	-0.0858**	-0.1182*
	(0.0183)	(0.0393)	(0.0268)	(0.0489)
Video	-0.0007	-0.0018	-0.0006	-0.0015
	(0.0159)	(0.0158)	(0.0159)	(0.0158)
Video × Foreign	0.0202	0.0222	0.0200	0.0217
	(0.0200)	(0.0200)	(0.0200)	(0.0199)
East	0.0539**			0.1247***
	(0.0180)			(0.0370)
East × Foreign	-0.0059			-0.0253
	(0.0225)			(0.0479)
Labor market tightness		-0.0723***		-0.0534**
		(0.0199)		(0.0204)
Labor market tightness × Foreign		0.0174		0.0131
		(0.0240)		(0.0246)
AfD votes			0.0016	-0.0077**
			(0.0015)	(0.0030)
AfD votes × Foreign			0.0002	0.0020
			(0.0018)	(0.0038)
N	3828	3828	3828	3828

Notes: Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, robust standard errors (in parentheses). Covariates included are layout, order, picture, name, firm size and educational demand. Labor market tightness is the relation of Applications to Positions in the labor market regions. Votes AfD are the first votes in percent in the electoral districts for the Alternative for Germany (AfD) in the Bundestag Election 2017.

applicants are available, with no evidence for ethnic discrimination depending on the regional and apprenticeship specific labor market tightness. The results for both, East Germany and the labor market tightness are robust as well as significant when considered simultaneously in column (6).

Educational demand and firm size

Two other interesting observations in Figure 4.2 result from firm size and educational requirements, both related to a firm's hiring procedures. We observe, consistent with Kaas and Manger (2012), that discrimination decreases with firm size. This is intuitive since increasing firm size leads to more professionalized screening procedures allowing us to solve asymmetric information to a higher degree. Secondly, we observe unequal treatment to be related to a firm's educational requirements with the highest discrimination by firms which have no educational requirements. To test the robustness of these patterns in a multivariate setting, Table 4.5 tests for heterogeneous treatment effects based on firm size and a firm's educational requirements. As before, we observe clear evidence for ethnic discrimination with the same effects of the video treatment for German- and Arabic-sounding names as in Table 4.2 for all specifications. When testing for the relevance of a firm's requirements, we observe a pattern consistent with Figure 4.2. With increasing requirements, the effect of ethnic discrimination decreases, but none of the interaction terms are statistically significant. When considering heterogeneity based on firm size, we are not able to observe the previous pattern or any significant difference. Neither of the results change in a joined analysis.

Time until the begin of the apprenticeship

To investigate the sensitivity of our results related to the timing of the applications, Table 4.A.5 tests for heterogeneity based on the time until the beginning of the apprenticeship and ethnicity and their interaction. There are two potential explanations for effect heterogeneity resulting from the timing of the application. On the one hand, employers searching for an apprentice shortly before the beginning of the apprenticeship have a smaller applicant pool that leads to the risk of not finding an apprentice. As a result, employers might discriminate less against minority groups given the lack of alternatives. On the other hand, with less remaining time until the start of an apprenticeship an application could signal a low level of effort in a previous job search and therefore a

Table 4.5: Probability of a callback of category C1: educational demand and firm size (OLS)

	(1)	(2)	(3)
Foreign	-0.1269**	-0.0999***	-0.1450**
	(0.0453)	(0.0269)	(0.0508)
Video	-0.0010	-0.0014	-0.0011
	(0.0148)	(0.0148)	(0.0148)
Video × Foreign	0.0211	0.0211	0.0211
	(0.0195)	(0.0195)	(0.0195)
Educational demand interaction (Reference category: None × Foreign):			
9 years × Foreign	0.0337		0.0346
	(0.0488)		(0.0487)
9 years+ × Foreign	0.0379		0.0364
	(0.0532)		(0.0528)
10 years × Foreign	0.0563		0.0550
	(0.0442)		(0.0439)
12/13 years × Foreign	0.0592		0.0586
	(0.0563)		(0.0562)
Firm size interaction (Reference category: Less than 6 × Foreign):			
6 to 50 × Foreign		0.0232	0.0211
		(0.0258)	(0.0258)
51 to 500 × Foreign		0.0297	0.0265
		(0.0286)	(0.0287)
501 or more × Foreign		0.0086	0.0060
		(0.0554)	(0.0548)
Educational demand	✓	✓	✓
Firm size	✓	✓	✓
N	3828	3828	3828

Notes: OLS. Callback category 1. Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, the cluster robust standard errors (in parentheses) are adjusted for clusters at the firm level. Dependent variable is callback category 1. Other covariates included: firm effects, educational demand, firm size, occupation, city, pictures, names, layout and order.

Table 4.6: Mean callbacks: schools

	Callback category 1		Callback category 2		
	Mean	Std. Dev.	Mean	Std. Dev.	N
School 1	0.090	0.286	0.153	0.360	1926
German 1	0.116	0.320	0.195	0.397	948
Foreign 1	0.064	0.246	0.112	0.316	978
School 2, German 2	0.131	0.338	0.222	0.416	962
School 3, Foreign 2	0.062	0.241	0.123	0.329	940

low level of motivation, as well as being a signal for low overall productivity given their previous failure to succeed in finding an apprenticeship. Results of Table 4.A.5 show that there is little evidence that these interaction terms drive the results in a certain direction. For callback category 1 the job start in 5 months and in 3 months interacted with ethnicity are not significantly different to the reference category (6 months). At the same time, the coefficients of the foreign and the video-foreign interaction terms stay almost unchanged, indicating that there are no heterogeneous treatment effects based on these variables.[10]

Schools

The four identities in the experiment are from three different schools. This is done in order to be able to test for differences between as well as within schools. However, it is possible that employers look up town areas where schools are located. As a consequence their decision as to whether or not to send an invitation might relate to implicit associations with these districts. To make sure the results are not influenced by this, we present callbacks for schools separately in Table 4.6. In callback category 1 the callback rate for Germans is very similar and is 11.6 and 13.1% respectively. The callback rates in the ethnic minority group are almost identical and are 6.4 and 6.2%. The same holds for the standard deviations. For callback category 2 relative differences between

[10]How long after the deadline candidates received callbacks did not differ for ethnicity due to a general drop in callbacks at the end of July. There were however some firms that invited candidates for internships again in the following year after their offers had been declined in 2018.

candidates are of a similar size. The school indicators are included in a regression with school number 1 being the base category. Results are depicted in Table 4.A.4. Results in (1) of Table 4.A.4 are similar to results already seen in the main estimation.[11] The coefficients of the second and third school show only a small and insignificant effect. While they are only slightly more popular among recruiters it is reassuring to see that the effect sizes in (1) of school 2 and school 3 are very similar. Again, for callback category 2 results show larger but still insignificant school effects.

4.5 Conclusion

Individuals without German citizenship have only half the apprenticeship entry rate compared to those with German citizenship (BMBF, 2018). This is explained with differences in occupational preferences, regional labor markets, limited proficiency in the German language and participation rate in internships. But what role does discrimination play? Given the importance of the school-to-work transition for future labor market prospects, it is important to understand the relevance of discrimination and to search for solutions in the school-to-work transition.

In this paper a correspondence experiment is used to study ethnic discrimination in the hiring market for apprenticeships in Germany. We send fictitious applications with German and Arabic-sounding names to real vacancies, augmenting applications with a short motivational video. These video introductions by applicants serve as a simple way of extending employer insight into the personality and motivation of applicants and therefore solving the problem of discrimination in the hiring market for apprenticeships.

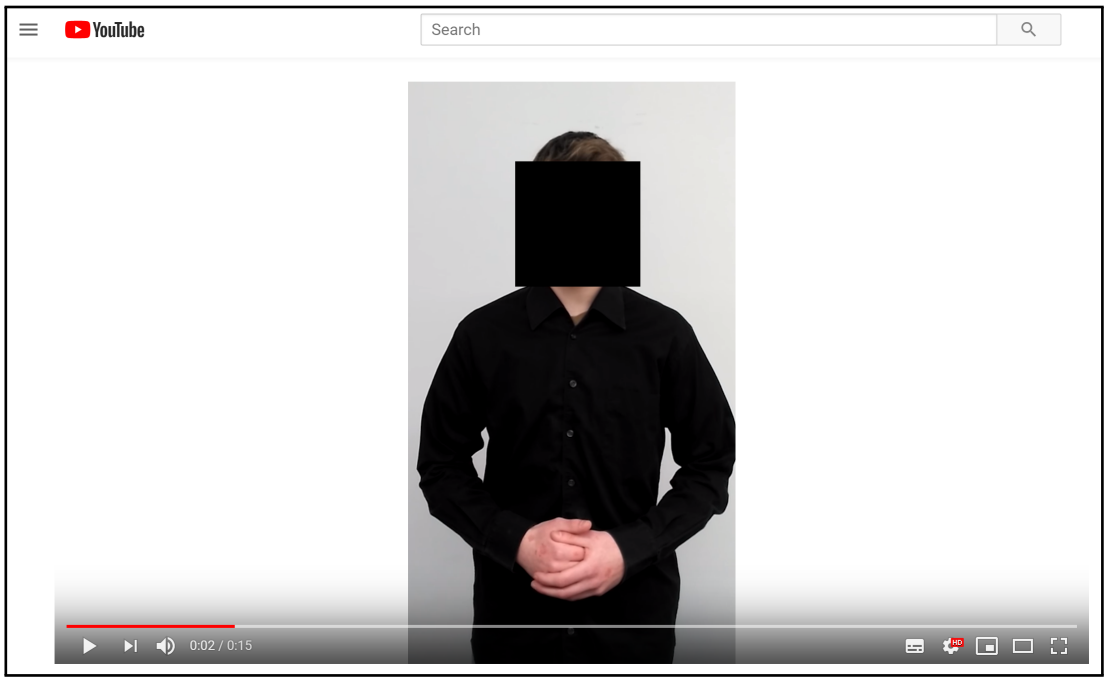
Previous experiments in German-speaking labor markets provide evidence for statistical discrimination (Kaas and Manger, 2012), taste-basted discrimination (Weichselbaumer, 2016a) as well as attention discrimination (Bartoš et al., 2016). When statistical discrimination is an important driver for ethnic discrimination in the labor market, one can expect higher levels of discrimination in the hiring market for apprenticeships compared to student internships (Kaas and Manger, 2012) and other labor markets where more work experience is preconditioned (e.g. Bertrand and Mullainathan, 2004).

[11] The variables for the candidates names are not included as a control in this model due to them being highly correlated with the school variables and causing collinearity.

Results show that applications with German-sounding names receive twice as many callbacks as those with Arabic-sounding names. This is in line with above mentioned literature which finds that individuals without German citizenship have only half the apprenticeship entry rate compared to those with German citizenship. Results suggest that most of the differences in apprenticeship entry rates can be attributed to ethnic discrimination. Providing a video in the application that gives employers more insight into the personality and motivation of applicants increases invitations to job interviews for Arabic-sounding names. However, since these conditional effects are insignificant we interpret the results as weak evidence of videos increasing invitations for Arabic-sounding names. Furthermore, despite the fact that videos only had observable effects on Arabic-sounding names, the videos of German-sounding names were viewed 37% more often compared to the Arabic-sounding names, which is in support of attention discrimination. This results in a dilemma that additional information as well as effort is not equally valued for all applicants. An observed unequal treatment related to a firm's educational requirements further supports this. Finally the analysis considers a firm's regional access to applicants for an apprenticeship as well as whether possible differences between East and West Germany are due to the recent success of a right-wing political party in East Germany. Our results show no evidence of discrimination depending on a firm's regional access to applicants. There is also no evidence of a higher degree of discrimination in East Germany.

4.A Appendix

Figure 4.A.1: Video screenshot



Note: Recruiters see video without censor bar.

Table 4.A.1: Summary: callback category 1

	German		Foreign		Difference		
	Mean	Std. Dev.	Mean	Std. Dev.	b	t-stat	Obs.
Callback category 1:							
Overall	0.124	0.329	0.063	0.243	0.060***	(6.46)	3828
Without video	0.125	0.330	0.056	0.231	0.068***	(6.04)	2552
With video	0.121	0.327	0.076	0.266	0.045**	(2.70)	1276
East Germany	0.162	0.369	0.097	0.296	0.066**	(3.25)	1098
West Germany	0.108	0.311	0.050	0.217	0.058***	(5.68)	2730
East Germany, without video	0.169	0.376	0.082	0.275	0.087***	(3.59)	732
East Germany, with video	0.148	0.356	0.126	0.332	0.022	(0.61)	366
Occupation:							
Automotive Technician	0.127	0.333	0.041	0.197	0.086***	(4.95)	1008
Sales	0.177	0.382	0.085	0.280	0.092***	(4.38)	1011
Office Clerk	0.091	0.288	0.064	0.245	0.027*	(2.16)	1809
Firm size:							
Less than 6	0.101	0.301	0.011	0.104	0.090***	(4.69)	552
6 to 50	0.127	0.333	0.058	0.235	0.068***	(5.41)	2079
51 to 500	0.114	0.319	0.082	0.274	0.032^{+}	(1.75)	1029
501 or more	0.220	0.416	0.174	0.382	0.045	(0.73)	168
Educational Demand:							
None (*Keine Anforderung*)	0.186	0.391	0.051	0.222	0.135**	(3.08)	219
9 years (*Hauptschulabschluss*)	0.147	0.355	0.080	0.272	0.067**	(2.76)	675
9 years+ (*Erw. Hauptschulabschluss*)	0.111	0.314	0.047	0.212	0.063*	(2.31)	381
10 years (*Mittlere Reife*)	0.117	0.321	0.062	0.240	0.055***	(4.80)	2436
12/13 years (*Hochschulreife*)	0.065	0.248	0.073	0.262	-0.008	(-0.17)	117
Month of job posting							
February	0.095	0.294	0.041	0.197	0.055***	(4.13)	1413
March	0.144	0.351	0.065	0.246	0.079***	(3.84)	870
April	0.141	0.349	0.079	0.271	0.062**	(2.82)	825
May	0.136	0.344	0.084	0.278	0.052*	(2.23)	720

Notes: All variables refer to callback category 1. Two-sample t-test for unpaired data and heteroscedasticity with null hypothesis of equal means performed. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4.A.2: Summary: callback category 2

	German		Foreign		Difference		
	Mean	Std. Dev.	Mean	Std. Dev.	b	t-stat	Obs.
Callback category 2:							
Overall	0.209	0.407	0.118	0.322	0.091***	(7.68)	3828
Without video	0.204	0.403	0.116	0.320	0.088***	(6.09)	2552
With video	0.219	0.414	0.121	0.327	0.098***	(4.68)	1276
East Germany	0.246	0.431	0.158	0.366	0.087***	(3.62)	1098
West Germany	0.194	0.396	0.102	0.302	0.092***	(6.86)	2730
East Germany, without video	0.243	0.430	0.153	0.360	0.090**	(3.08)	732
East Germany, with video	0.251	0.435	0.169	0.376	0.082^{+}	(1.93)	366
Occupation:							
Automotive Technician	0.198	0.399	0.095	0.293	0.103***	(4.67)	1008
Sales	0.295	0.456	0.177	0.382	0.118***	(4.47)	1011
Office Clerk	0.165	0.372	0.099	0.299	0.066***	(4.18)	1809
Firm size:							
Less than 6	0.144	0.352	0.047	0.213	0.096***	(3.90)	552
6 to 50	0.219	0.413	0.106	0.308	0.112***	(7.02)	2079
51 to 500	0.205	0.404	0.144	0.352	0.061**	(2.59)	1029
501 or more	0.329	0.473	0.326	0.471	0.004	(0.05)	168
Educational Demand:							
None (*Keine Anforderung*)	0.265	0.443	0.128	0.336	0.137*	(2.54)	219
9 years (*Hauptschulabschluss*)	0.242	0.429	0.152	0.359	0.090**	(2.96)	675
9 years+ (*Erw. Hauptschulabschluss*)	0.232	0.423	0.115	0.320	0.116**	(3.03)	381
10 years (*Mittlere Reife*)	0.197	0.398	0.108	0.311	0.089***	(6.14)	2436
12/13 years (*Hochschulreife*)	0.097	0.298	0.109	0.315	-0.012	(-0.22)	117
Month of job posting							
February	0.160	0.367	0.087	0.282	0.073***	(4.23)	1413
March	0.237	0.426	0.116	0.320	0.122***	(4.77)	870
April	0.247	0.432	0.131	0.338	0.116***	(4.28)	825
May	0.230	0.422	0.163	0.370	0.067*	(2.27)	720

Notes: All variables refer to callback category 1. Two-sample t-test for unpaired data and heteroscedasticity with null hypothesis of equal means performed. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4.A.3: Probability of a callback of category C2 (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Foreign	-0.0911***	-0.0916***	-0.0983***	-0.0991***	-0.0996***	-0.0990***	-0.0993***	-0.0882***	-0.0964***	-0.0914***
	(0.0119)	(0.0119)	(0.0221)	(0.0219)	(0.0218)	(0.0216)	(0.0216)	(0.0201)	(0.0226)	(0.0208)
Video	0.0105	0.0106	0.0118	0.0118	0.0119	0.0119	0.0119	0.0119	0.0161	0.0054
	(0.0127)	(0.0127)	(0.0126)	(0.0126)	(0.0125)	(0.0124)	(0.0124)	(0.0112)	(0.0196)	(0.0179)
Video × Foreign									-0.0084	0.0130
									(0.0250)	(0.0247)
Layout, Order		✓	✓	✓	✓	✓	✓	✓	✓	✓
Picture, Name			✓	✓	✓	✓	✓	✓	✓	✓
Firm size				✓	✓	✓	✓		✓	
City					✓	✓	✓		✓	
Occupation						✓	✓		✓	
Educational Demand							✓		✓	
HR Gender: female, unknown							✓		✓	
Firm FE								✓		✓
N	3828	3828	3828	3828	3828	3828	3828	3828	3828	3828
adj. R^2	0.015	0.017	0.019	0.033	0.046	0.063	0.065	0.502	0.065	0.502

Notes: OLS. Callback category 2. Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, robust standard errors (in parentheses) for (1)-(7) and (9), cluster robust standard errors adjusted for clusters at the firm level for (8) and (10).

Table 4.A.4: Probability of a callback: schools (OLS)

	(1) Callback category 1	(2) Callback category 2
Foreign	-0.0853***	-0.1066***
	(0.0164)	(0.0199)
Video	-0.0014	0.0054
	(0.0148)	(0.0179)
Video × Foreign	0.0211	0.0130
	(0.0195)	(0.0247)
School 2	0.0085	0.0226
	(0.0132)	(0.0162)
School 3	0.0065	0.0151
	(0.0107)	(0.0147)
N	3828	3828

Notes: Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, the cluster robust standard errors (in parentheses) are adjusted for clusters at the firm level. Other variables included are: firm effects, pictures, layout and order.

Table 4.A.5: Probability of a callback: job start in months (OLS)

	(1) Callback category 1	(2) Callback category 2
Foreign	-0.0834***	-0.1155***
	(0.0246)	(0.0286)
Video	-0.0015	0.0048
	(0.0148)	(0.0178)
Video × Foreign	0.0215	0.0141
	(0.0195)	(0.0248)
Job Start interaction (Ref. category: Job Start in 6 months × Foreign):		
Job Start in 5 months × Foreign	-0.0012	0.0069
	(0.0260)	(0.0305)
Job Start in 4 months × Foreign	-0.0169	-0.0099
	(0.0302)	(0.0356)
Job Start in 3 months × Foreign	-0.0010	0.0081
	(0.0329)	(0.0394)
N	3828	3828

Notes: Significance levels $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, the cluster robust standard errors (in parentheses) are adjusted for clusters at the firm level. Other variables included are: firm effects, job start in months (levels), pictures, layout and order.

Bibliography

Altonji, J. G. and C. R. Pierret (2001). Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Angrist, J. D. and A. B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Antoni, M., N. Bachbauer, J. Eberle, M. Antoni, N. Bachbauer, J. Eberle, and B. Vicari (2018). NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7515). FDZ Datenreport 02/2018. Technical report, Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Arrow, K. (1973). The theory of discrimination. *Discrimination in labor markets 3*(10), 3–33.

Ashenfelter, O. and A. Krueger (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *The American Economic Review 84*(5), 1157–1173.

Attar, I. and D. Cohen-Zada (2018). The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. *Journal of Economic Behavior & Organization 153*, 38–57.

Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics 21*(1), 1–42.

BIBLIOGRAPHY

Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005 BT - Audit Studies: Behind the Scenes with Theory, Method, and Nuance. pp. 63–77. Cham: Springer International Publishing.

Baert, S., B. Cockx, N. Gheyle, and C. Vandamme (2015). Is There Less Discrimination in Occupations Where Recruitment Is Difficult? *ILR Review 68*(3), 467–500.

Bahrs, M. and M. Schumann (2016). Unlucky to Be Young? The Long-Term Effects of School Starting Age on Smoking Behaviour and Health. Hche research paper, Hamburg.

Bartnitzky, H. and R. Christiani (1977). *Zeugnis ohne Zensuren: Hilfen für die Praxis des Zeugnisschreibens*. Bagel.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016). Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review 106*(6), 1437–1475.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago and London: University of Chicago Press.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy 70*(5, Part 2), 9–49.

Becker, G. S. (1964). *Human capital*. Columbia University Press, New York.

Bedard, K. and E. Dhuey (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics 121*(4), 1437–1472.

Beicht, U. (2016). Ausbildungschancen von Ausbildungsstellenbewerbern undbewerberinnen mit Migrationshintergrund. *Aktuelle Situation*.

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

Bénabou, R. and J. Tirole (2002). Self-Confidence and Personal Motivation*. *The Quarterly Journal of Economics 117*(3), 871–915.

Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination. In A. V. B. Experiments and E. D. B. T. H. of Economic Field (Eds.), *Handbook of Field Experiments*, Volume Volume 1, pp. 309–393. North-Holland.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review 94*(4), 991–1013.

BIBB (2018). Ergebnis der Erhebung des Bundesinstituts für Berufsbildung (BIBB) über neu abgeschlossene Ausbildungsverträge zum 30. September 2017.

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *The American Educator 25*(1), 24–28.

Biewen, M. and M. Tapalaga (2017). Life-Cycle Educational Choices in a System with Early Tracking and 'Second Chance' Options. *Economics of Education Review 56*, 80–94.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *Review of Economics and Statistics 93*(2), 455–467.

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

BMBF (2018). Berufsbildungsbericht 2018. *Bundesministerium für Bildung und Forschung (BMBF)*.

Bönke, T., G. Corneo, and H. Lüthen (2015). Lifetime Earnings Inequality in Germany. *Journal of Labor Economics 33*(1), 171–208.

Bos, W., S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, and R. Valtin (2007). *No Title*. Münster: Waxmann Verlag.

BIBLIOGRAPHY

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic policy 22*(52), 782–861.

Bryson, A. and A. Chevalier (2015). Is there a taste for racial discrimination amongst employers? *Labour Economics 34*, 51–63.

Buckles, K. S. and D. M. Hungerman (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *The Review of Economics and Statistics 95*(3), 711–724.

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

Card, D. (1999). The Causal Effect of Education on Earnings. *Handbook of Labor Economics 3*, 1801–1863.

Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Carlsson, M., G. B. Dahl, B. Öckert, and D.-O. Rooth (2014). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics 97*(3), 533–547.

Carlsson, M., L. Fumarco, and D.-O. Rooth (2018). Ethnic discrimination in hiring, labour market tightness and the business cycle-evidence from field experiments. *Applied Economics 50*(24), 2652–2663.

Carlsson, M. and D. O. Rooth (2012). Revealing taste-based discrimination in hiring: A correspondence testing experiment with geographic variation. *Applied Economics Letters 19*(18), 1861–1864.

Carnevale, A. P., S. J. Rose, and B. Cheah (2011). The college payoff: Education, occupations, lifetime earnings. Technical report, Georgetown University Center on Education and the Workforce.

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

Cascio, E. U. and D. W. Schanzenbach (2016). First in the Class? Age and the Education Production Function. *Education Finance and Policy 11*(3), 225–250.

Chetty, R., M. Stepner, S. Abraham, and E. Al (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA 315*(16), 1750–1766.

Chevalier, A., S. Gibbons, A. Thorpe, M. Snell, and S. Hoskins (2009). Students' Academic Self-Perception. *Economics of Education Review 28*(6), 716–727.

Chowdry, H., C. Crawford, and A. Goodman (2011). The Role of Attitudes and Behaviours in Explaining Socio-Economic Differences in Attainment at Age 16. *Longitudinal and Life Course Studies 2*(1), 59–76.

Cook, P. J. and S. Kang (2016). Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation. *American Economic Journal: Applied Economics 8*(1), 33–57.

Cook, P. J. and S. Kang (2018). The School-Entry-Age Rule Affects Redshirting Patterns and Resulting Disparities in Achievement. *NBER Working Paper Series* (24402).

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review 25*(1), 43–62.

Day, J. C. and E. C. Newburger (2002). The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. Technical report, U.S. Census Bureau.

Deming, D. and S. Dynarski (2008). The Lengthening of Childhood. *Journal of Economic Perspectives 22*(3), 71–92.

Dhuey, E., D. Figlio, K. Karbownik, and J. Roth (2017). School Starting Age and Cognitive Development.

Dickson, M. and C. Harmon (2011). Economic Returns to Education: What We Know, What We Don't Know, and Where We Are Going-Some Brief Pointers. *Economics of Education Review 30*(6), 1118–1122.

Dillon, E. W. (2018). Risk and Return Trade-Offs in Lifetime Earnings. *Journal of Labor Economics 36*(4), 981–1021.

Dobkin, C. and F. Ferreira (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review 29*(1), 40–54.

Doleac, J. L. and L. C. D. Stein (2013). The Visible Hand: Race and Online Market Outcomes. *The Economic Journal 123*(572), F469–F492.

Dubey, P. and J. Geanakoplos (2010). Grading exams: 100,99,98,... or A,B,C? *Games and Economic Behavior 69*(1), 72–94.

Dustmann, C. (2004). Parental Background, Secondary School Track Choice, and Wages. *Oxford Economic Papers 56*(2), 209–230.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009). Revisiting the German Wage Structure *. *Quarterly Journal of Economics 124*(2), 843–881.

Dustmann, C., P. A. Puhani, and U. Schönberg (2017). The Long-term Effects of Early Track Choice. *The Economic Journal 127*, 1348–1380.

Elder, T. E. and D. H. Lubotsky (2009). Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers. *Journal of Human Resources 44*(3), 641–683.

Ewens, M., B. Tomlin, and L. C. Wang (2014). Statistical Discrimination or Prejudice? A Large Sample Field Experiment. *The Review of Economics and Statistics 96*(1), 119–134.

Fertig, M. and K. Görlitz (2013). Missing Wages: How to Test for Biased Estimates in Wage Functions? *Economics Letters 118*(2), 269–271.

Figlio, D. N. (2005). Names, expectations and the black-white test score gap. *NBER Working Paper Series* (11195), 1–31.

Filippin, A. and M. Paccagnella (2012). Family Background, Self-Confidence and Economic Outcomes. *Economics of Education Review 31*(5), 824–834.

Fitzenberger, B. (1999). *Wages and Employment Across Skill Groups: An Analysis for West Germany*. Heidelberg: ZEW Economic Studies Physika-Verlag.

Fitzenberger, B., A. Osikominu, and R. Völter (2005). Imputation Rules to Improve the Education Variable in the IAB Employment Subsample.

Fletcher, J. and T. Kim (2016). The effects of changes in kindergarten entry age policies on educational achievement. *Economics of Education Review 50*, 45–62.

Fredriksson, P. and B. Öckert (2006). Is early learning really more productive? The effect of school starting age on school and labor market performance.

Fredriksson, P. and B. Öckert (2014). Life-cycle Effects of Age at School Start. *The Economic Journal 124*(579), 977–1004.

Gartner, H. (2005). The imputation of wages above the contribution limit with the German IAB employment sample. *FDZ-Methodenreport 2*(2).

Gehrer, K., S. Zimmermann, C. Artelt, and S. Weinert (2012). The Assessment of Reading Competence. Technical report, University of Bamberg, National Educational Panel Study, Bamberg.

Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*, 1219–1260.

Goldberg, A., D. Mourinho, and U. Kulke (1996). Labour market discrimination against foreign workers in Germany. *International Migration Papers 7*.

Goldin, C. and C. Rouse (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review 90*(4), 715–741.

Gregg, P. and E. Washbrook (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 11. *Longitudinal and Life Course Studies 2*(1), 41–58.

Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2016). General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle. *Journal of Human Resources*.

Hanushek, E. A. and L. Wößmann (2006). Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries*. *The Economic Journal 116*(510), C63–C76.

Hartenstein, M. and G. H. Ruddies (1978). *Der Schulbericht: Arbeitsanleitung für Grundschullehrer; Verstehenshilfe für Eltern.* Calwer Verlag.

Hastings, J. S. and L. Tejeda-Ashton (2008). Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico. *NBER Working Paper Series* (14538).

Hayes, D. P. and M. G. Ahrens (1988). Vocabulary simplification for children: a special case of 'motherese'? *Journal of Child Language 15*(2), 395–410.

Heckman, J. J., J. E. Humphries, S. Urzua, and G. Veramendi (2011). The Effects of Educational Choices on Labor Market, Health, and Social Outcomes. *Unpublished manuscript, University of Chicago, Department of Economics*.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. *Handbook of the Economics of Education 1*, 307–458.

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

Helbig, M. and R. Nikolai (2015). *Die Unvergleichbaren: Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949.* Julius Klinkhardt.

Heller, K. A. and H. Bartnitzky (1984). Leistungsdiagnostik in der Schule.

Hurwitz, M., J. Smith, and J. S. Howell (2015). Student Age and the Collegiate Pathway. *Journal of Policy Analysis and Management 34*(1), 59–84.

Ichino, A. and R. Winter-Ebmer (1999). Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments. *European Economic Review 43*(4-6), 889–901.

Jachmann, M. (2003). *Noten oder Berichte?* VS Verlag für Sozialwissenschaften.

Jähnen, S. and M. Helbig (2015). Der Einfluss Schulrechtlicher Reformen auf Bildungsungleichheiten Zwischen den Deutschen Bundesländern. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie 67*(3), 539–571.

Jibuti, D. (2018). Discrimination against Workers with Visible Tattoos: Experimental Evidence from Germany. *CERGE-EI Working Paper Series* (628).

Johnson, R. W. and A. G. Feng (2013). Financial Consequences of Long-Term Unemployment during the Great Recession and Recovery. *Urban Institute: Unemployment and Recovery Project 13*.

Julian, T. and R. Kominski (2011). Education and Synthetic Work-Life Earnings Estimates. American Community Survey Reports. ACS-14. *US Census Bureau*.

Kaas, L. and C. Manger (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review 13*(1), 1–20.

Kahneman, D. and A. Deaton (2010). High Income Improves Evaluation of Life but not Emotional Well-Being. *Proceedings of the National Academy of Sciences 107*(38), 16489 LP – 16493.

KMK (1960). Kultusministerkonferenz Beschluss vom 8./9.12.1960.

KMK (1970). Kultusministerkonferenz: Empfehlungen zur Arbeit in der Grundschule.

Koch, A., J. Nafziger, and H. S. Nielsen (2015). Behavioral economics of education. *Journal of Economic Behavior & Organization 115*(Supplement C), 3–17.

Koppensteiner, M. F. (2018). Relative Age, Class Assignment, and Academic Performance: Evidence from Brazilian Primary Schools. *The Scandinavian Journal of Economics 120*(1), 296–325.

Krause, A., U. Rinne, and K. F. Zimmermann (2012). Anonymous job applications in Europe. *IZA Journal of European Labor Studies 1*(1), 5.

Kruppe, T. and S. Unger (2014). Effictiveness of Data Correction Rules in Process-Produced Data. *IAB-Discussion Paper 15*(15).

Kübler, D., J. Schmid, and R. Stüber (2019). Take Your Time to Grow: A Field Experiment on the Hiring of Youths. *German Economic Review 0*(0).

Lahey, J. N. and D. R. Oxley (2018). Discrimination at the intersection of age, race, and gender: Evidence from a lab-in-the-field experiment. *NBER Working Paper Series* (25357).

Landersø, R., H. S. Nielsen, and M. Simonsen (2017). School Starting Age and the Crime-age Profile. *Economic Journal 127*(602), 1096–1118.

Lange, S. and M. von Werder (2014). The effects of delayed tracking: Evidence from German states. Courant Research Centre: Poverty, Equity and Growth - Discussion Papers 163, Göttingen.

Larsen, E. R. and I. F. Solli (2017). Born to run behind? Persisting birth month effects on earnings. *Labour Economics 46*, 200–210.

Lefgren, L. and F. McIntyre (2006). The Relationship between Women's Education and Marriage Outcomes. *Journal of Labor Economics 24*(4), 787–830.

Leigh, A. and C. Ryan (2008). Estimating returns to education using different natural experiment techniques. *Economics of Education Review 27*(2), 149–160.

List, J. A. (2004). The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field*. *The Quarterly Journal of Economics 119*(1), 49–89.

Lubotsky, D. and R. Kaestner (2016). Do 'Skills Beget Skills'? Evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood. *Economics of Education Review 53*, 194–206.

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being*. *The Quarterly Journal of Economics 120*(3), 963–1002.

Marsh, H. W. and J. W. Parker (1984). Determinants of Student Self-Concept: Is it Better to Be a Relatively Large Fish in a Small Pond Even if you Don't Learn to Swim as Well? *Journal of Personality and Social Psychology 47*(1), 213–231.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

McCrary, J. and H. Royer (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. *American Economic Review 101*(1), 158–195.

McEwan, P. J. and J. S. Shapiro (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *Journal of Human Resources 43*(1), 1–29.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy 66*(4), 281–302.

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

Mora, R. and I. Reggio (2015, oct). Didq: A Command for Treatment-effect Estimation under Alternative Assumptions. *The Stata Journal 15*(3), 796–808.

Murphy, K. M. and F. Welch (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics 8*(2), 202–229.

Murphy, R. and F. Weinhardt (2016). Top of the Class: the Importance of Ordinal Rank. *Working Paper* (February), 1–61.

Nam, K. (2014). Until when does the effect of age on academic achievement persist? Evidence from Korean data. *Economics of Education Review 40*, 106–122.

Nunley, J. M., M. F. Owens, and R. S. Howard (2011). The effects of information and competition on racial discrimination: Evidence from a field experiment. *Journal of Economic Behavior & Organization 80*(3), 670–679.

Nüß, P. (2018). Duration dependence as an unemployment stigma: Evidence from a field experiment in Germany. Economics Working Paper 2018-06, Kiel.

Oreopoulos, P. (2011). Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. *American Economic Journal: Economic Policy 3*(4), 148–171.

Pekrun, R. (1996). Ziffernzensuren oder Berichtszeugnisse? Drei Kritische Anmerkungen zur Annahme unterschiedlicher Wirkungen. *Benner/Merkens/Schmidt: Bildung und Schule im Tansformationsprozeß von SBZ, DDR und neuen Ländern-Untersuchungen zu Kontinuität und Wandel. Berlin: Freie Universität Berlin, Institut für Allgemeine Pädagogik*, 253–259.

Pellizzari, M. and F. C. Billari (2012). The younger, the better? Age-related differences in academic performance at university. *Journal of Population Economics 25*(2), 697–739.

Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review 56*, 152–176.

PISA (2007). *Die Ergebnissse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

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

Ponzo, M. and V. Scoppa (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling 36*(3), 578–599.

Protsch, P. and H. Solga (2015). How employers use signals of cognitive and noncognitive skills at labour market entry: Insights from field experiments. *European Sociological Review 31*(5), 521–532.

Psacharopoulos, G. (2014). *The Returns to Investment in Higher Education*, pp. 121–148. Rotterdam: Sense Publishers.

Puhani, P. A. and A. M. Weber (2007). Does the early bird catch the worm? *Empirical Economics 32*(2), 359–386.

Reichelt, M. (2015). Using Longitudinal Wage Information in Linked Data Sets. The Example of ALWA-ADIAB. *FDZ-Methodenreport 1*(1).

Riach, P. A. and R. Judith (2006). An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market. *Advances in Economic Analysis & Policy 6*(2).

Riach, P. A. and J. Rich (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal 112*(483), F480–F518.

Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. *IZA Discussion Paper Series* (8584).

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics 17*(3), 523–534.

Rose, H. and J. R. Betts (2004). The Effect of High School Courses on Earnings. *The Review of Economics and Statistics 86*(2), 497–513.

Rosenthal, R. and L. Jacobson (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* Holt, Rinehart & Winston.

Sakshaug, J. W. and F. Kreuter (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods 6*(2), 113–122.

Schmillen, A. and H. Stüber (2014). Lebensverdienste nach Qualifikation: Bildung lohnt sich ein Leben lang. Technical Report 1, IAB, Nürnberg.

Schnittjer, I. and C. Durchhardt (2015). Mathematical Competence: Framework and Exemplary Test Items. Technical report, University of Bamberg, National Educational Panel Study, Bamberg.

Schultz, T. W. (1960). Capital Formation by Education. *Journal of Political Economy 68*(6), 571–583.

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review 51*(1), 1–17.

Skirbekk, V., H.-P. Kohler, and A. Prskawetz (2004). Birth month, school graduation, and the timing of births and marriages. *Demography 41*(3), 547–568.

Skopek, J. (2013). Data Manual, Starting Cohort 6, release 3.0.1. Technical report, Bamberg: National Educational Panel Study (NEPS).

Smith, J. (2009). Can regression discontinuity help answer an age-old question in education? The effect of age on elementary and secondary school achievement. *The B.E. Journal of Economic Analysis & Policy 9*(1), 1–30.

Solga, H. and B. Kohlrausch (2013). How Low-achieving German Youth Beat the Odds and Gain Access to Vocational Training—Insights from Within-Group Variation. *European Sociological Review 29*(5), 1068–1082.

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics 45*(1), 523–538.

Statistisches Bundesamt (2017). Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2015/16. Fachserie 11, Reihe 1. Wiesbaden.

Statistisches Bundesamt (2018). Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2017/18. Fachserie 11, Reihe 1. Wiesbaden.

Steiner, V. and K. Wagner (1998). Has Earnings Inequality in Germany Changed in the 1980's? *Zeitschrift für Wirtschafts- und Sozialwissenschaften 118*, 29–59.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown and M. E. Curtis (Eds.), *The nature of vocabulary acquisition*, Chapter 6, pp. 89–106. Hillsdale, N.J.: Erlbaum.

Tamborini, C. R., H. M. Iams, and K. Whitman (2009). Marital History, Race, and Social Security Spouse and Widow Benefit Eligibility in the United States. *Research on Aging 31*(5), 577–605.

Tamborini, C. R., C. Kim, and A. Sakamoto (2015). Education and Lifetime Earnings in the United States. *Demography 52*(4), 1383–1407.

Tamm, M. (2013). The Impact of a Large Parental Leave Benefit Reform on the Timing of Birth around the Day of Implementation*. *Oxford Bulletin of Economics and Statistics 75*(4), 585–601.

Tjaden, J. D., C. Schwemmer, and M. Khadjavi (2018). Ride with Me—Ethnic Discrimination, Social Markets, and the Sharing Economy. *European Sociological Review 34*(4), 418–432.

Ullrich, H. and M. Wöbcke (1981). *Notenelend in der Grundschule: Alternative Beurteilungsformen für die Praxis*. Kösel.

Valfort, M.-A. (2017). Has France a problem with Muslims? Evidence from a field experiment in the labour market. *Working Paper*.

van Aken, M. A. G., A. Helmke, and W. Schneider (1997). Selbstkonzept und Leistung–Dynamik ihres Zusammenspiels: Ergebnisse aus dem SCHOLASTIK-Projekt. *Entwicklung im Grundschulalter*, 341–350.

Veit, S. and R. Yemane (2018). The ADIS study: A large-scale correspondence test on labor market discrimination in Germany - Technical Report. WZB Discussion Paper SP VI 2018-103, Berlin.

Wagner, C. and R. Valtin (2003). Noten oder Verbalbeurteilungen? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie 35*(1), 27–36.

Webber, D. A. (2016). Are College Costs Worth it? How Ability, Major, and Debt Affect the Returns to Schooling. *Economics of Education Review 53*, 296–310.

Weichselbaumer, D. (2016a). Discrimination Against Female Migrants Wearing Headscarves. *IZA Discussion Paper No. 10217*.

Weichselbaumer, D. (2016b). Discrimination Against Migrant Job Applicants in Austria: An Experimental Study. *German Economic Review 18*(2), 237–265.

Weinert, S., C. Artelt, M. Prenzel, M. Senkbeil, T. Ehmke, and C. H. Carstensen (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft 14*(2), 67–86.

Zeinz, H. and O. Köller (2006). Noten, soziale Vergleiche und Selbstkonzepte in der Grundschule. In A. Schründer-Lenzen (Ed.), *Risikofaktoren kindlicher Entwicklung: Migration, Leistungsangst und Schulübergang*, pp. 177–190. Wiesbaden: VS Verlag für Sozialwissenschaften.

Zschirnt, E. and D. Ruedin (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies 42*(7), 1115–1134.