ifo BEITRÄGE
zur Wirtschaftsforschung

105
2024

Improving Educational Outcomes: Analyses of Intervention and Public Opinion

Katharina Wedel

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

ifo BEITRÄGE zur Wirtschaftsforschung

105
2024

Improving Educational Outcomes: Analyses of Interventions and Public Opinion

Katharina Wedel

Herausgeber der Reihe: Clemens Fuest
Schriftleitung: Chang Woon Nam

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

Bibliografische Information der Deutschen Nationalbibliothek

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

ISBN Nr. 978-3-95942-131-7

Druck: Pinsker Druck und Medien GmbH, Mainburg

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

Improving Educational Outcomes: Analyses of Interventions and Public Opinion

Inaugural-Dissertation

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

eingereicht an der
Ludwig-Maximilians-Universität München
2023

vorgelegt von

Katharina Wedel

Referent: Prof. Dr. Ludger Wößmann
Korreferent: Prof. Dr. Joachim Winter
Promotionsabschlussberatung: 31.01.2024

Datum der mündlichen Prüfung:	09.01.2024
Namen der Berichterstatter:	Prof. Dr. Ludger Wößmann
	Prof. Dr. Joachim Winter
	Prof. Philipp Lergetporer, PhD

Preface

Katharina Wedel prepared this study while she was working at the Center for the Economics of Education at the ifo Institute. The study was completed in September 2023 and accepted as doctoral thesis by the Department of Economics at LMU Munich. It consists of four distinct empirical essays that address various aspects of the economics of education. Chapter 2 sheds light on the interaction of two inputs into the education production function: instruction time and teacher qualifications. The results show that teacher qualifications play a moderating role for the effect of instruction time on student achievement. Chapter 3 examines dropout from a mentoring program designed to help disadvantaged adolescents and analyzes a program agency's cost-benefit trade-offs in the decision to target additional interventions to prevent dropout. Chapter 4 investigates public opinion towards targeted financial support and the role of external circumstances compared to own effort for (educational) success. This chapter shows how information on the differences in academic-track attendance by parental background in Germany increases the perception that external circumstances determine educational success, and private donations to charities but does not affect demand for redistributive education spending. Finally, chapter 5 studies the consequences of technological change on individuals' labor-market expectations and their intentions to participate in further training. Experimental results show that information about the automatability of one's occupation affects labor-market expectations and increases the likelihood to participate in further training and retraining.

Keywords: Education Production Function, Instruction Time, Student Achievement, Teacher Qualifications, Program Dropout, Prediction, Machine Learning, Information, Survey Experiment, Charitable Donations, Policy Preferences, Automation, Further Training, Labor-Market Expectations
JEL-No: C21, C53, D04, D63, D64, D83, H11, H52, I21, I24, I25, I29, J24, O33

Acknowledgement

First and foremost, I would like to thank my supervisor Ludger Wößmann for his guidance and advice throughout these past years of preparing my dissertation. I am extremely grateful for his support and feedback as well as the opportunity to be part of his team with excellent research opportunities. Furthermore, I am very grateful to Joachim Winter for his valuable feedback and helpful discussions about the dissertation projects.

In addition, I want to thank my co-authors from the ifo Education Survey projects: Elisabeth Grewenig, Philipp Lergetporer, and Katharina Werner. Thank you for the great collaboration, the countless discussions and meetings. I also thank my co-author Sven Resnjanskij for his support, advice and discussions, not only in common projects, but also in my single-authored project. I have learned a lot from all of you.

I am grateful to all my current and former colleagues at the ifo Center for the Economics of Education for the great research environment, helpful discussions and feedback: Benjamin Arold, Alexander Bertermann, Raphael Brade, Vera Freundl, Sarah Gust, Franziska Hampf, Sarah Kersten, Lavinia Kinne, Lukas Mergele, Simon ter Meulen, Caterina Pavese, Franziska Pfaehler, Marc Piopiunik, Pietro Sancassani, Florian Schoner, Moritz Seebacher, and Larissa Zierow. Having them as colleagues has truly been a pleasure. Connections with some of them have extended beyond the workplace, forming lasting friendships. Also special thanks go to Ulrike Baldi-Cohrs and Franziska Binder for their support in many areas that go beyond administrative assistance at ifo. Further, I want to thank my peers and friends from other ifo centers as well as from the LMU Munich, who accompanied my academic journey - thank you for sharing the PhD experience, including seminars, conferences, and all the ups and downs.

I would also like to thank Jan Bietenbeck for sparking my interest in the economics of education and for providing helpful feedback on the second chapter of this dissertation. I am also grateful to Simon Wiederhold for valuable discussions about some chapters of this dissertation.

During fall 2022, I had the incredible opportunity to visit the Program on Education Policy and Government (PEPG) at Harvard University. I am most grateful to the whole team of PEPG for this opportunity. I have tremendously benefited from numerous discussions, conversations and seminar participations. I am also thankful for the generous financial support from the CESifo Young Ambassador Program and the CRC Rationality & Competition for this research stay.

Last, but not least, I would like to express my deepest gratitude to my husband, my family and friends for their unconditional love and support – thank you for always believing in me.

Katharina Wedel

Contents

List of Figures

List of Tables

1 Introduction

1.1 Economics of Education: Returns, Production, and Public Opinion

Education stands as a cornerstone for the future prosperity of economies, not only providing individuals a path to individual labor-market success and well-being, but also functioning as a crucial determinant of economic growth on a larger scale. This recognition of education's pivotal role in shaping both individual outcomes and broader economic prosperity has led to a surge in scholarly attention on the economics of education (Hanushek et al., 2016).

The concept of *human capital*, viewed as the stock of skills and knowledge embodied in an individual determining productivity, has gained remarkable significance in the discourse surrounding economic growth. Scholars such as Hanushek and Woessmann (2012, 2015) have underscored the connection between human capital and long-term economic advancement. In the realm of the macroeconomics literature, the determinants of economic growth have been subject to extensive exploration. Various theoretical growth models have emphasized distinct mechanisms by which education influences economic growth. The most basic model, the Solow-Swan growth model by Solow (1956) and Swan (1956), relates economic output to capital and labor. Building upon this model, the augmented neoclassical growth model by Mankiw et al. (1992) incorporated human capital as an integral factor for economic growth: in this model, human capital increases through accumulation of education, thereby raising the steady-state level of income. Different from this model are the endogenous growth models: in these models, human capital increases the innovative capacity of the economy and generates economic growth through the creation of new technologies (Lucas, 1988; Romer, 1990; Howitt and Aghion, 1998).

At the individual level, the literature has shown the profound influence of higher human capital on improved outcomes such as higher wages and lower unemployment. This exploration traces its origins back to the 1950s to the seminal contributions of Mincer (1958), Schultz (1961), and Becker (1962), containing theoretical and empirical work. They formalized individuals' cost-benefit considerations with respect to their educational investment decisions as a trade-off between the benefits and the direct as well as indirect costs of education. The costs consist of monetary costs, such as tuition fees or schooling material, and time or opportunity costs, such as foregone earnings. Their groundbreaking work laid the foundation for a vast realm of research on education within the field of economics, setting the stage for the formulation of the human capital theory, which culminated in Becker's pioneering work in 1964.

The far-reaching impacts of education span not only improved pecuniary outcomes such as higher wages, but also non-pecuniary outcomes and outcomes unrelated to the labor market, such as health outcomes. Many studies, such as Card (1999) and Heckman et al. (2006), have substantiated the influence of individuals' human capital on wages. Beyond financial metrics, this impact extends to aspects such as unemployment, where research by Ashenfelter and Ham (1979) and Nickell (1979) has demonstrated the substantial role played by human capital. Similarly, the realm of health has not remained untouched, with research by Cutler and Lleras-Muney (2006) and Lochner (2011), among others. In addition, an increase in educational achievement has been shown to lead to a decrease in future occurrences of violent and property crimes (Lochner, 2020).

For a long time, human capital was measured by the average years of schooling in empirical work. However, this measure assumes that, regardless of the education system, a year of schooling yields the same increase in skills and knowledge (Hanushek and Woessmann, 2021). Recent advancements have shifted the focus from mere years of education to more nuanced metrics, such as student cognitive skills measured through standardized achievement tests in mathematics, science and reading (see e.g., Hanushek and Woessmann, 2011).

Closely related is the literature on educational production, which posits that certain inputs are related to skill and human capital through a production function, as established by Hanushek (1986). This line of literature focuses more on the determinants and sources of human capital and skills. Among the commonly studied inputs are family attributes, individual ability, and the quality and quantity of school inputs, such as school resources and teacher quality (Hanushek, 2020). The usual outcome is student achievement, frequently measured by students' test scores. At the institutional level, the educational production is influenced by factors like tracking (Hanushek and Woessmann, 2006), school accountability mechanisms, such as standardized testing with external comparison compared to internal testing (Bergbauer et al., 2021), expenditures (Jackson et al., 2016, 2021), and preferences (Figlio et al., 2019; Hanushek et al., 2022). Similarly, at the school level, numerous factors come into play, each with the potential to influence students' educational outcomes. Factors such as class size (Woessmann and West, 2006; Angrist et al., 2019), teachers (Chetty et al., 2014; Hanushek et al., 2019), teaching methodologies (Schwerdt and Wuppermann, 2011; Bietenbeck, 2014), and instruction time (Lavy, 2015; Rivkin and Schiman, 2015) have gathered extensive attention in the literature. Chapter 2 of this dissertation adds to this literature by examining the interaction between two of these factors: instruction time and teacher qualifications. In this chapter, I find a positive impact of instruction time on students' test scores across all countries, which is significantly larger for students with better qualified teachers. Investigating effect heterogeneity across countries, I find that instruction time has no significant effect in developing countries on average, but it increases students' test scores when taught by a highly qualified teacher also in these countries. Several developing countries have already expanded schooling opportunities but often the majority of students still does not reach the threshold of basic literacy of 400 test score points (Hanushek and Woessmann, 2015). In lower-middle-income countries and low-income countries, for example, the share of students not reaching basic skill levels is 81.3 and 90.5 percent, respectively (Gust et al., 2022). Thus, developing countries still lag behind developed countries in terms of student achievement. Finding ways to improve student achievement in developing countries therefore helps address inequalities between countries or regions.

However, not only institutional factors and resources play a role in determining students' outcomes. Educational success, especially in Germany, depends to a large extent also on family background, such as parental socio-economic status (Schütz et al., 2008; Woessmann et al., 2023). This link results in a concerning persistence of unequal educational outcomes, which inadvertently reinforces preexisting societal inequalities. In response, researchers, policy-makers, and other agents such as non-governmental organizations, are prompted to explore avenues that can alleviate the stark inequality prevalent in educational attainment.

A growing part of the literature focuses on interventions offered through (non-profit or non-governmental) organizations that aim at improving the outcomes of disadvantaged children outside the school context, often in the form of tutoring or mentoring programs. Examples are, among others, the "Pathways to Education" program offered by the *Regent Park Community Health Centre*, a non-profit, community-based organization in Toronto, the "Becoming A Man" program by the Chicago non-profit *Youth Guidance*, or the mentoring program by the international social franchise *Rock Your Life!*. These organizations aim not only for an improvement in student achievement but also for an enrichment in students' social environment through the provision of social programs and interventions that provide disadvantaged children with better opportunities. Benefits from such program attendance for the participating children have been established in the literature (Rodríguez-Planas, 2012; Heller et al., 2017; Oreopoulos et al., 2017; Kosse et al., 2020; Resnjanskij et al., 2024). In addition to the participants' perspective, it also provides valuable insights to focus on a program agency's perspective. As these social programs are frequently publicly funded, it is important to use the resources in the best possible and most effective way which is why identification of students at risk of dropout becomes crucial. Chapter 3 adds to this literature by examining a program agency's optimal behavior for targeting additional interventions to counteract the threat of dropout from a social program, such as a mentoring program, that is designed to help disadvantaged children. Preventing dropout from social programs is important to preclude adverse consequences for participants, such as negative emotions, lower self-esteem, a higher likelihood to skip school, and increased alcohol consumption. Often, the most disadvantaged students who would benefit most from participation have a higher dropout risk which makes it crucial to think about interventions that prevent dropout.

Furthermore, governments often provide education and influence student outcomes by setting the institutional structure of a country's education system as well as the allocation of resources to schools. These processes are intricately linked to the political process: the outcome of democratic elections profoundly influences the way educational resources are distributed and institutional structures are shaped. This underlines the central role played by public opinion in determining the trajectory of education policies. Understanding the factors that drive public opinion regarding education becomes also important in addressing inequalities. Eliciting public opinion not only sheds light on the prevailing attitudes of the population but also potentially informs about public support for governmental strategies aimed at mitigating inequalities related to students' socio-economic status. By eliciting the preferences of the public, a deeper understanding can be gained regarding the factors that contribute to these opinions. Thus, the interconnectedness of educational outcomes, political systems, and public opinion is of high importance. It not only underscores the need to understand and address the mechanisms perpetuating educational inequalities but also highlights the significance of public perceptions in shaping educational policies and efforts to rectify existing inequalities. Chapter 4 adds to this topic by analyzing how providing information about differences in academic-track attendance rates between students from more and less advantaged backgrounds in Germany might have implications for public approval of targeted financial support. We do this by examining individuals' donation decisions to charities supporting students from disadvantaged backgrounds as well as support for governmental redistributive education spending as outcome measures.

Finally, inequalities do not only appear in childhood and adolescence, but also in adulthood. Due to the rise of automation, a pressing concern emerges, for example, regarding the unequal impact of technological and structural change on different parts of the population: the demand for skills differentially changes due to technological advancements (Dauth et al., 2021; OECD, 2021). This inequality prompts crucial questions about the government's potential interventions to counteract the unequal consequences on labor-market outcomes. Since researchers and policymakers argue that further training for adults is a potential strategy to empower individuals with skills relevant to the changing labor-market requirements and to mitigate the impacts of technological change, it is important to understand the public's attitudes and opinion regarding their perceived vulnerability to automation, the drivers of further training participation and their support for policies targeted at those most affected by technological and structural change. Chapter 5 adds to this topic by examining these questions. Understanding the public's opinion not only informs policy decisions but also holds the potential to shape effective strategies that alleviate inequalities arising from the influence of technological change and automation.

The remainder of the introduction of this dissertation is structured as follows: section 1.2 provides an overview of the data used in this dissertation. I use international and national survey data for the analyses. Section 1.3 describes the empirical methods used (randomization, fixed effects estimation and novel Machine Learning (ML) methods). Section 1.4 concludes by providing an overview of the four chapters that constitute this dissertation and the policy implications that can be drawn from these.

1.2 Data

This section describes the data sources that I use in this dissertation: the Trends in International Mathematics and Science Study (section 1.2.1), the data from the evaluation of a mentoring program in Germany (section 1.2.2), and the ifo Education Survey (section 1.2.3).

1.2.1 Trends in International Mathematics and Science Study

Chapter 2 uses data from the *Trends in International Mathematics and Science Study* (TIMSS) which is a data source for international student assessments. TIMSS conducts international large-scale assessments of students' skills in mathematics and science and is administered every four years since 1995. Participating countries are, for example, developing countries (such as Chile, Oman, and Saudi Arabia) but also developed countries (such as France, the United States of America, and Japan). TIMSS samples entire classes and provides students' test scores as well as information on student, teacher, and school characteristics gained through linked questionnaires from students, parents, teachers, and school principals (Martin et al., 2016; TIMSS, 2019).

1.2.2 Data from a Mentoring Program

Chapter 3 uses data from an evaluation of a mentoring program called *Rock Your Life!* in Germany (see Resnjanskij et al., 2024). This mentoring program assigns students from eighth and ninth grade in lowest-track secondary schools in Germany (mentees) to university students (mentors). The evaluation study consists of students from rather disadvantaged backgrounds who were assigned a mentor, i.e., to the treatment group, but also of students who were part of a control group and thus were not assigned a mentor. Chapter 3 uses the data on students who were assigned a mentor only. The data comprises information on the students' family background and their household composition, information about school-related topics, such as grades and school hours, as well as personal characteristics, such as the Big Five personality traits, and information about the mentoring relationship, such as whether the relationship is still active one year after the program start.

1.2.3 ifo Education Survey

Chapters 4 and 5 use data from the *ifo Education Surveys* 2019 and 2022, respectively. The ifo Education Survey is an annual, representative opinion survey in Germany conducted since 2014 (Freundl et al., 2023). It is a repeated cross-section of participants aged 18 and older. The sample is drawn according to quotas to be representative of the German adult population. In recent years, respondents were sampled through an online platform, and they answered the questions online on a personal device. In addition to opinion questions about topics related to education, the survey collects a variety of respondents' background characteristics. Often, the survey includes randomization of respondents in treatment and control groups, where the treatment group is provided with additional information while the control group is not (see also section 1.3.1 for an overview of survey experiments). Comprehensive explanations of specific attributes of the distinct survey waves can be found within the corresponding chapters (chapters 4 and 5).

1.3 Empirical Methods

This section describes the empirical methods used in this dissertation. First, I will discuss survey experiments (section 1.3.1). Second, I will describe the use of student fixed effects as an identification strategy since exploiting random assignment to groups is not always possible (section 1.3.2). Third, I will briefly mention ML methods (section 1.3.3).

1.3.1 Survey Experiments

The use of *survey experiments* has become popular among economists over the past years (Haaland et al., 2023). The concept behind these experiments involves randomly selecting one or several subsets of participants to receive varying versions of the same question. More specifically, chapters 4 and 5 use information provision experiments in online surveys, in which a randomly chosen subgroup is provided with information while the other subgroup is not. Thus, these experiments are particularly suited to understanding how providing information to participants influences their beliefs, preferences, and decision-making processes. Since outside a controlled setting information acquisition is likely endogenous, the random assignment of information to a treatment group provides a good solution to construct counterfactuals and cleanly identify causal effects of information provision.

Survey experiments are basically field experiments, recording respondents' stated preferences and usually observing participants in their natural environment. One must acknowledge that biases might arise from non-random selection into participation in an online survey. However, participation costs in (online) survey experiments are low because travel costs to laboratories as in lab experiments are not an issue since respondents can participate using their personal devices. Thus, we can draw from a quite diverse pool of participants. In addition, when conducting the survey, we also make sure that the sample reflects the German population using quotas for gender, age, state, education level, and employment status. Furthermore, Grewenig et al. (2023) find that when using re-weighted non-probabilistic internet surveys, the response patterns closely resemble those of mixed-mode surveys, both statistically and quantitatively.

1.3.2 Student Fixed Effects Model

To ascertain the causal effect of instruction time on student achievement which is the aim in chapter 2, it is necessary to have exogenous variation that remains uncontaminated by any hidden disparities among schools and students. Just examining a simple correlation between instruction time and student achievement is not suited to estimate the causal effect of instruction time on student achievement because it is unlikely that students with different background characteristics and ability are randomly assigned to classes or subjects with different instruction time. For example, if students who tend to perform better in a specific subject are assigned more instruction time in that specific subject, the estimated relationship will be upward-biased. Controlling for observable characteristics, such as students' socio-economic background or their gender, is also unlikely to yield causal estimates. Unobservable characteristics, such as student ability, might still bias the estimates if they affect student achievement and if they are correlated with instruction time. As a solution, I follow the literature that addresses concerns of unobservable characteristics and student sorting (e.g., Metzler and Woessmann, 2012; Bietenbeck et al., 2018) and use a *student fixed effects model*. With that model, I exploit within-student between-subject variation which controls for unobservable characteristics such as student ability. The appeal of this model stems from the fact that students have the same general skill level in two subjects and are exposed to the same school environment in both subjects.

1.3.3 Machine Learning Methods

ML methods are increasingly used in the economics literature (Currie et al., 2020). Chapter 3 and part of the analyses in chapter 4 also use ML methods. While ML is a vast field, I limit the overview in this section to the methods used in this dissertation.

In chapter 3, the main aim is to generate predictions of an outcome variable using many input variables. To do so, the chapter mainly uses two supervised ML methods: the *Least Absolute Shrinkage and Selection Operator* (LASSO) and the *Random Forest*. These two models are used to learn the relationship between a range of information about adolescents participating in a mentoring program on the one hand, such as information about the family background and the household composition, information about school-related topics, as well as personal characteristics, and dropout from the mentoring program on the other hand. These predictions from ML methods are then used in a model set up to describe a program agency's decision to target additional interventions to reduce dropout. Detailed descriptions about the methods can be found in chapter 3.

Chapter 4 uses causal ML. More specifically, it uses a *Causal Forest* algorithm building on the Random Forest to estimate Conditional Average Treatment Effects (CATEs) which are the predicted treatment effects for out-of-sample observations. This data-driven approach allows the researcher to detect heterogeneous effects in subgroups and to capture high-dimensional, more complex combinations of covariates that might have been missed otherwise.

1.4 Chapter Overview

This section provides a short summary of the four essays that are part of this dissertation and delves into the resulting policy implications. Each essay corresponds to one chapter, is self-contained, and can be read independently. Chapter 2 sheds light on the interaction of instruction time and teacher qualifications, two inputs into the education production function. Chapter 3 examines dropout from a mentoring program designed to help disadvantaged adolescents and analyzes a program agency's cost-benefit trade-offs in the decision to target additional interventions to prevent dropout. Chapter 4 investigates public opinion towards targeted financial support and the role of external circumstances compared to own effort for (educational) success. Chapter 5 studies the consequences of technological change on individuals' labor-market expectations and their intentions to participate in further training.

Chapter 2 investigates two inputs to the education production function: instruction time and teachers. Thus, this chapter focuses on the interaction between the quantity and the quality of instruction. I use international TIMSS data from 2015, as described in section 1.2.1, and exploit within-student between-subject variation by implementing a student fixed effects model, as described in section 1.3.2. I find that on average, an additional hour of instruction time has a positive impact on students' test scores across all countries. Importantly, these effects of instruction time are significantly larger for students with better-qualified teachers. Teacher qualifications are measured through participation in professional development in the relevant subject, possessing a Bachelor's degree (or higher) with the relevant subject as the major subject, teacher education with a specialization in the subject, as well as teaching experience (in years). While instruction time has no significant effect in developing countries on average, it increases students' test scores when taught by a highly qualified teacher also in developing countries.

Findings from prior literature and the findings of chapter 2 highlight that instruction time stands as an important factor in enhancing student achievement. Notably, the qualifications of teachers significantly influence the impact of instruction time on students' academic outcomes. This positive synergy between instruction time and teacher qualifications also holds importance for policy considerations: the simplicity of slightly extending instruction time underscores its feasibility for implementation, offering a straightforward avenue to enhance student achievement. However, given the financial implications associated with additional instruction time, policymakers must ascertain its cost-effectiveness. The outcomes of chapter 2 suggest that the link between instruction time and teacher quality is integral to student achievement and that teacher qualifications should be considered as well, especially in developing countries. This insight underscores the essential interplay between these factors and emphasizes their combined influence on student success.

Chapter 3 analyzes dropout from a mentoring program from a program agency's perspective. This project is joint work with Sven Resnjanskij. Using ML algorithms (see section 1.3.3), we predict participants' dropout risk. We find that important variables for predicting dropout are adolescents' family environment, the student's math performance, and personal characteristics such as self-efficacy and engagement in extracurricular activities. Through a model analyzing the cost-benefit considerations of the agency, we highlight how a program agency can make use of dropout risk predictions and can optimally respond to the threat of program dropout. We demonstrate that even algorithms with relatively low predictive power can still result in a considerable increase in expected program returns.

Understanding dropout from an agency's perspective is of high importance given the typically large scale and frequent public funding associated with mentoring programs. This emphasizes the need to maximize the use of these resources, while also deploying efficient interventions that reduce the risk of dropout after enrollment, making it essential to identify students at risk of dropping out. Moreover, the significance of preventing dropout extends beyond individual programs; it has far-reaching implications for the aggregate sum of foregone benefits. Consequently, establishing strategies and interventions aimed at mitigating and preventing dropout becomes a crucial agenda for both policymakers and researchers alike to help disadvantaged adolescents.

Chapter 4 examines how perceptions of differences in education outcomes of students from more and less advantaged parental backgrounds shape preferences for private donations and for redistributive education spending as well as respondents' view of the role of external circumstances and effort for educational success. This chapter is joint work with Elisabeth Grewenig and Katharina Werner. In this chapter, we use a survey experiment (see section 1.3.1) among the German adult population and find that information about the correlation of parental background and academic-track attendance of students strongly increases the perception that external circumstances rather than effort determine educational success. These effects persist in a follow-up survey conducted two weeks later. Importantly, information provision also significantly increases private donations to charities supporting students from disadvantaged socio-economic backgrounds but does not affect the demand for redistributive education spending by the government. This pattern of results may be explained by differences in the perceived transparency regarding the opportunity costs of funds used in both spending decisions.

Our findings show that the provision of information about prevailing inequalities does not inherently fail to evoke specific support for individuals from disadvantaged backgrounds. Instead, the support appears to decline when asked about preferences for governmental redistribution in terms of educational expenditures. Nevertheless, the outcomes shed light on a distinct trend – respondents display a genuine concern regarding the educational differences among students from different socio-economic backgrounds, as evidenced by variations in academic-track attendance rates. This concern translates into a willingness to extend assistance through charitable donations. Consequently, even if respondents' attitude towards advocating increased redistributive education spending remains unaltered through information provision, the act of informing them about inequalities, particularly linked to parental backgrounds, emerges as a promising avenue for raising their involvement and engagement.

Chapter 5 investigates how beliefs about the automatability of one's occupation affect individuals' labor-market expectations and their willingness to participate in further training. This chapter is joint work with Philipp Lergetporer and Katharina Werner. This chapter uses data from the ifo Education Survey 2022 (see section 1.2.3). In the online survey, we find that respondents on average underestimate the automatability of their occupation, especially those in high-automatability occupations (i.e., those in occupations with an automatability above 50 percent). In a survey experiment (see section 1.3.1), we randomly provide a treatment group with information about the automatability of respondents' occupations which increases their concern about the future and expectation of changes in their work environment. More importantly, the information increases respondents' willingness to participate in further training, especially among workers in high-automatability occupations by five percentage points, nearly closing the gap to those in occupations with low automatability.

The findings of chapter 5 underscore the potential benefits of disseminating information regarding the automatability of various occupations to the public. Such efforts could effectively enhance awareness regarding the potential consequences of technological advancements on the labor market, particularly for individuals working in occupations highly susceptible to automation. This, in turn, could stimulate a forward-looking response by motivating individuals to engage in further training initiatives. By proactively preparing themselves for the evolving demands accelerated by technological shifts, individuals could position themselves more effectively within changing requirements.

2 Instruction Time and Student Achievement: The Moderating Role of Teacher Qualifications*

2.1 Introduction

Quantity and quality of instruction are essential for students' educational achievement. On the one hand, instructional quantity has a positive impact on a student's achievement (e.g., Lavy, 2015; Rivkin and Schiman, 2015). On the other hand, instructional quality has proven to be important for student achievement (e.g., Rockoff, 2004; Hanushek and Rivkin, 2006). Hence, there may be complementarities between the quality and quantity of instruction: more instruction time will probably be of even larger benefit if teachers use the additional time efficiently, e.g., by covering new or revising old content instead of using the time for classroom management or administrative tasks. Thus, more instruction time leads to better performance if teachers actively use the time for teaching. Furthermore, it is important how well a teacher knows the subject and how well she is able to explain it to her students (Carroll, 1989). Patall et al. (2010) also state that "the effectiveness of instruction" (p. 430) can influence whether additional time at school affects students' outcomes, and if so, whether the effect is positive or negative.

The effect of instruction time might go in different directions. On the one hand, additional instruction time in a subject might give the teacher the opportunity to cover more material, analyze and discuss it in more detail, take the time to answer students' questions, and combine concepts that arise in different classes (National Center on Time & Learning, 2017). On the other hand, students might get to a point where more instruction time and thus more input is harmful since they are unable to absorb further information. Some pupils might even get bored, especially the already high-performing students. Andersen et al. (2016) argue that students need to be motivated to follow and focus on what they are being taught to achieve long-term success so that they can benefit from the extra instruction time. The authors explain that students need self-control to focus. However, students' ability to concentrate decreases with more instruction time, making it more difficult for them to pay attention and control their emotions and thoughts. Ultimately, students might become less focused and even more aggressive (Andersen et al., 2016).

* This chapter is based on the paper: Wedel, K. 'Instruction time and student achievement: The moderating role of teacher qualifications.' *Economics of Education Review*, 85, 102183, 2021. https://doi.org/10.1016/j.econedurev.2021.102183.

In this paper, I study to what extent the effect of instruction time on student performance is moderated by the quality of teachers. I add instructional time as an input to the education production function and interact its effect with teacher qualifications. I use a student fixed effects model which accounts for observable and unobservable individual-specific factors, such as unobserved ability, and exploit within-student between-subject variation to identify the effect of an additional hour of instruction time on test scores.

I use data from the 2015 Trends in International Mathematics and Science Study (TIMSS) which contain two observations for each student, one in math and one in science. My main outcome variable is students' test scores in math and science as a measure of their cognitive skills. The main independent variable is the instruction time in these subjects, measured in hours per week. Instruction time is defined as the "amount of time during which students receive instruction from a classroom teacher in a school [...] context" (UNESCO, 2023). It does not include teacher training days, holidays, breaks at school, or learning time outside of school, such as time for homework and tutoring.

My results show that teacher qualifications play a moderating role for the effect of instruction time on student achievement. In the student fixed effects model, I regress student test scores on instruction time as well as on teacher qualifications and their interactions. On average, across all countries, one hour more instruction time leads to 0.030 standard deviations higher test scores, with boys (0.037 standard deviations) benefiting more than girls (0.024 standard deviations). It is important to note that the effect varies according to teachers' formal qualifications: it is larger for students with better qualified teachers. On average, the effect of instruction time on student achievement by a highly qualified teacher is 0.042 to 0.058 standard deviations, depending on which teacher qualification is considered. These teacher qualifications are measured in terms of participation in professional development, a Bachelor's degree (or higher) with the relevant subject as a major subject as well as teacher education with a specialization in the relevant subject.

The moderating role of teacher quality is particularly telling in understanding the effect of instruction time in developing countries. As in a previous analysis of Programme for International Student Assessment (PISA) data by Lavy (2015), I find that the effect of instruction time is larger for students in developed countries than in developing countries. In fact, for an extended sample of developing countries covered by TIMSS, including countries in the Middle East, the average effect of instruction time in developing countries is not statistically significant and close to zero. However, also in developing countries, instruction time by a highly qualified teacher increases test scores by 0.016 to 0.027 standard deviations.

I verify the robustness of my results using a series of varying specifications. First, I apply a within-teacher specification to rule out bias due to unobserved teacher characteristics. The results are robust: the coefficients are slightly larger than those in the main specification. Furthermore, the results are robust to including a squared term of instruction time, as well as to restricting the sample to schools that do not track students into different classes according to their ability.

In addition to test scores, I also examine whether more instruction time in a subject affects students' motivation and attitude towards that subject. On the one hand, additional instruction time in a subject can cause a student to become tired of that subject which does not necessarily lead to lower test scores but instead students may develop an aversion to the subject. On the other hand, more instruction time may encourage students to enjoy the subject even more since they are able to go into more detail about specific content. My results suggest that more instruction time leads to a more positive attitude towards the subject. However, this attitude effect is not influenced by the qualifications of teachers.

This paper is organized as follows: section 2.2 presents the related literature, followed by a description of the data in section 2.3 and of the empirical strategy in section 2.4. The empirical results are presented in section 2.5. Section 2.6 concludes.

2.2 Literature

This paper lies at the intersection of two strands of the literature: the one on the impact of instruction time on student achievement and the one on teacher quality and qualifications and how these relate to a student's achievement.

Recent literature on the impact of instruction time on students' test scores mostly finds a positive impact. This paper mainly relates to two recent articles which use international data. First, Lavy (2015) examines the effect of instruction time in 50 different countries using the 2006 PISA study, which evaluates students in math, science, and reading. Applying student fixed effects, Lavy (2015) finds that an increase in instruction time of one hour on average leads to 0.06 standard deviations higher test scores.[1] Second, Rivkin and Schiman (2015) use 2009 PISA data and apply a student fixed effects model that exploits variation in instruction time within schools across subjects, or across grades. They find that increasing the weekly instruction time by one hour leads to test scores that are 0.02 to 0.03 standard deviations higher. In addition, they show that better classroom environment in terms of student behavior and teacher-student interaction enhances the positive effect of additional instruction time (Rivkin and Schiman, 2015).

Other studies use national data. Bingley et al. (2018) apply the same methodology as Lavy (2015) and Rivkin and Schiman (2015) to Danish administrative data. Bingley et al. (2018),

[1] Cattaneo et al. (2017) replicate Lavy's (2015) study, applying the same methodology to PISA data for Switzerland. Their focus on only one country has the advantage of allowing them to examine the effects of distinct uses of instruction time in a context with very similar curricula and educational objectives which do not cause biases (Cattaneo et al., 2017). They also deviate from the approach employed by Lavy (2015) by using official teaching times recommended by the education authorities instead of data reported by students.

in comparison, use data for three cohorts of students in each grade throughout their entire compulsory education. Hence, they focus on the accumulated time from one grade to the next. These effects are about twice as large as those based on the time of only one grade.

Further studies focus on the length of the school year. Some of these use exogenous variation in instruction time due to school closures or weather-related absences as an instrument for instruction time (Marcotte, 2007; Goodman, 2014) and find adverse effects of a lack of instruction time on student achievement. Besides, Aucejo and Romano (2016) compare the effectiveness of an extension of the school year by ten days to a reduction in absences by a similar amount and show that the latter results in higher gains in terms of test scores in math and reading.

Several other studies examine the length of the school day. Figlio et al. (2018) document positive effects from the provision of additional instruction time in literacy on students' reading scores in Florida. Similarly, Battistin and Meroni (2016) find positive effects on math test scores due to an expansion of math and reading instruction in lower secondary schools in Italy. Meroni and Abbiati (2016) exploit the same expansion in Italy and show that girls' attitude towards math increases while boys' attitude decreases due to increased instruction time in that subject. In addition, Bellei (2009) studies the transition from part-time to full-time school days in Chile. His findings also suggest a positive impact on student achievement in math and reading. Further studies exploit school reforms, for example in Germany, which led to an increase in the number of weekly hours of instruction in academic-track high schools (Dahmann, 2017; Huebener et al., 2017). These authors find small but positive effects of instructional hours. Andersen et al. (2016) use a randomized controlled trial (RCT) in which some schools were randomly assigned more instruction time. They find that a student's reading performance increases by 0.15 standard deviations due to an increase in instruction time. In addition, Wu (2020) disentangles the effects of instruction time in hours per day and days per year to focus on the distribution of instructional hours. He finds that it is rather the length of the school day that drives the positive effect on student achievement than the length of the school year.

The second strand of the literature to which this paper refers is concerned with the impact of teacher quality and qualifications on student achievement. Teacher quality is found to be a major driver and determinant of student achievement and has often been measured in terms of teacher value-added (e.g., Hanushek, 1971; Rivkin et al., 2005; Koedel et al., 2015), assessing the quality of a teacher in terms of the gain in student achievement from one year to another. Chetty et al. (2014), for example, find that an increase in teacher value-added by one standard deviation results in 0.14 standard deviations higher student achievement in math. Even though Hanushek and Rivkin (2010) state that experience and education only "explain little of the variation in teacher quality" (p. 267), other studies have focused on teacher qualifications, such as certification and college major, an approach that I follow due to data constraints.[2] The evidence on the relationship between student achievement and these teacher qualifications is rather mixed as I will present below: one part of the literature finds no returns to better qualified teachers (e.g., Hanushek, 1986; Rivkin et al., 2005) while others find positive effects on student achievement (e.g., Goldhaber and Brewer, 2000; Clotfelter et al., 2007).[3] However, in the public debate and politics, certain teacher qualifications are required in recruitment processes and play a major role in compensation and tenure decisions (e.g., Podgursky and Springer, 2007; Shuls and Trivitt, 2015). In the U.S., for example, the No Child Left Behind act required that all core subject matter teachers are highly qualified, which means that they should hold a Bachelor's degree, should be certified or licensed by the state, and should be able to demonstrate subject matter competence (Jacob, 2007).

Advanced degrees, teacher licensures, and certifications are often considered as teacher qualifications that might affect student performance. Several studies find insignificant or even negative associations (e.g., Buddin and Zamarro, 2009; Sass, 2015). Buddin and Zamarro (2009), for example, examine the impact of performance in teacher licensure tests and more traditional measures on student achievement in elementary schools in California. Their results suggest that holding an advanced degree or a teacher certification is unrelated to student achievement. Kane et al. (2008) also find little difference in student test performance by initial certification status of their teachers. However, other studies report positive effects: Goldhaber and Brewer (1997, 2000) point out the relevance of subject-specific degrees. Goldhaber and Brewer (2000) find that students who are taught by teachers with a Bachelor's or Master's degree in math tend to have higher test scores in math compared to those students who are taught by teachers with an out-of-subject degree. Hence, having a teacher with subject-specific preparation, that is with a degree or certification in the subject, can result in higher achievement for students compared to being taught by a teacher without subject-specific training (see also Shuls and Trivitt, 2015). In addition, Clotfelter et al. (2007) and Goldhaber and Anthony (2007) show clear evidence that having a teacher who is board certified by the National Board for Professional Training Standards leads to higher student achievement.

Studies on subject-matter expertise as well as a teacher's major in the subject that she teaches consistently find positive associations: Rockoff et al. (2011) examine new teachers in New York City and look at non-traditional predictors of effectiveness, such as content knowledge and personality traits. As a proxy for content knowledge, they use the number of courses taken in a subject or the college major, finding larger gains for students in terms of test scores with teachers who majored in science or math. Besides, Angrist and Guryan (2008) measure teacher quality by the teachers' educational background, such as whether the teacher majored in the subject she teaches. Metzler and Woessmann (2012) use subject-specific teacher test

[2] To measure teacher value-added, one needs at least two observations per student at two points in time, ideally one at the beginning of a school year and one at the end. However, TIMSS assesses each student at only one point in time: students are assessed only once in fourth grade, but in the two subjects math and science. Therefore, I cannot use a value-added approach or include a lagged dependent variable in the model.

[3] Manning et al. (2019) provide evidence in a meta-analysis that higher qualifications of teachers are associated with higher classroom quality and higher quality early childhood education and care environments.

scores to assess a teacher's subject-matter knowledge and show a positive impact on student achievement in math. Similarly, Bietenbeck et al. (2018) find a modest impact of teacher subject knowledge on student performance.

Several studies analyze teaching experience, of which some find small to no returns to more experienced teachers (Hanushek, 1986, 1997; Buddin and Zamarro, 2009). In contrast, according to Clotfelter et al. (2007), more experienced teachers are better at raising student achievement than inexperienced teachers. Other studies also document such positive returns to experience (Wiswall, 2013; Papay and Kraft, 2015). Especially during the first few years of their career, teachers tend to perform worse than teachers with more experience (Rivkin et al., 2005; Staiger and Rockoff, 2010; Harris and Sass, 2011).

Lee and Lee (2020) combine all three qualifications (experience, advanced degrees and subject knowledge) into an index and examine the impact of highly qualified teachers on student achievement in a cumulative perspective: they show that being taught by several highly qualified teachers results in acquiring higher academic degrees.

Participation in professional development and teacher training has also been evaluated as a teacher qualification that matters for student achievement, with most of the studies on the U.S. finding no effects (Jacob and Lefgren, 2004; Garet et al., 2010). In contrast, Angrist and Lavy (2001) find positive impacts of teacher training in Jerusalem elementary schools on students' test scores similar to Lucas et al. (2014) in developing countries.

To the best of my knowledge, none of these previous studies addresses the complementarities between the quantity and quality of instruction. Hence, I contribute to the literature by exploring the interaction between instruction time and teacher quality, measured by the formal qualifications of teachers. Using a new data source, the 2015 TIMSS data, allows me to show results for developing countries that were not considered in previous studies on instruction time. These are especially countries from the Middle-East, such as Saudi Arabia, the United Arab Emirates, and Oman, as well as Singapore and Kazakhstan. Hence, I present new evidence on the smaller effect of instruction time in developing countries, and how teacher qualifications play a moderating role for this effect. Moreover, I focus on fourth-graders, which is especially relevant since young children are particularly sensitive to interventions and the return on investment in human capital is higher (Cunha et al., 2006). In addition, I examine not only the impact on cognitive skills, but also on students' motivation and their attitude towards the subject.

2.3 Data

2.3.1 Trends in International Mathematics and Science Study

I use data from the Trends in International Mathematics and Science Study (TIMSS) which are a repeated cross-section. The TIMSS & PIRLS International Study Center at the International Association for the Evaluation of Educational Achievement (IEA) conducts assessments of students' achievements in math, science and reading, which are internationally comparable. In general, this study is conducted in more than 60 countries (TIMSS, 2019). In addition to information about a student's achievement, the data also include information about students' attitudes, teachers, school resources, and instructional practices (TIMSS, 2019). This information is collected in separate questionnaires: there is a student questionnaire, a home questionnaire which parents are asked to fill out, a teacher questionnaire, and a school and curriculum questionnaire that is filled out by the school principal.

TIMSS uses a two-stage random sample design: in the first stage, a sample of schools is determined, and in the second stage, one or more classes within a school are selected (Martin et al., 2016). Since TIMSS monitors students' instructional and curricular experiences at the classroom level, TIMSS samples classes rather than individual students (Martin et al., 2016). TIMSS monitors student achievement along with the other previously mentioned information for two grades: fourth and eighth grade. The national samples of students in TIMSS are constructed to describe the target population. Theoretically, the two-stage random sample design generates samples of students with the same probability of selection (Martin et al., 2016). In practice, however, a varying number of selected classes and differential non-response can result in different probabilities of selection, requiring individual sampling weights for the students. I use the senate weights which can be used when differences across countries are examined. In this way, each country receives the same weight.

Since the pool of questions and items in TIMSS is too large to be answered by one student, students receive only a subset of questions (a so-called booklet) to answer. Each booklet contains both math and science questions. This is different from the PISA data where some students only answer questions in some subject areas, but not in all three. In PISA data, students were assigned scores according to their performance in other domains (Jerrim et al., 2017). According to Jerrim et al. (2017), the results could then be driven by a random imputation error. However, my results are not affected by this since all students answer both math and science questions. Aggregating the results of all booklets yields results for the overall assessment. Then, plausible values are imputed to obtain five estimates of a student's achievement (Martin et al., 2016).

2.3.2 Analysis Sample and Variables

I use data on fourth grade students from the 2015 survey wave since interventions matter more for younger students.[4] The final sample includes 108,358 students in 1,586 classes and 4,283 schools in 39 countries. Since every student is evaluated twice – once in math, once in science – the number of observations amounts to 216,716. I present descriptive statistics in Table 2.1.

The dependent variable is the *test score* of a student in math or science. The test score variable in the data set is the plausible value for math and science.[5] I standardize the test scores so that the mean is zero and the standard deviation is one. These test scores measure a student's cognitive attainment in math and science (Woessmann, 2003).

As an alternative dependent variable, I generate an index of four variables for each subject to obtain a measure of a student's motivation and attitude towards a subject to evaluate whether additional instruction time affects a student's motivation and attitude. Questions are for example "I enjoy learning mathematics" and "I learn many interesting things in mathematics" as well as the corresponding questions for science. The students could answer on a four-point scale which ranges from "agree a lot" to "disagree a lot".[6] I use factor analysis estimating polychoric correlations which account for the categorical nature of the underlying variables to generate the index *like subject*. To do so, I first obtain the polychoric correlations matrix of the four variables which I then use to perform the factor analysis. One factor is retained according to the Kaiser criterion which indicates keeping factors with eigenvalues larger than one (Backhaus et al., 2011). The factor loadings and scoring coefficients are reported in Appendix Table A2.1. To assess the robustness, I generate an alternative index by following the procedure described by Kling et al. (2007): I first normalize each variable by subtracting the mean and dividing by the standard deviation. I then add up the normalized versions of the four variables and take the average, that is, I divide by four which is the number of variables.[7]

[4] In the analysis I use only those countries for which information on test scores and instruction time in both math and science is available and where science is taught as a separate subject. Science comprises life science, physical science, and earth science. In general, math test scores are available for all countries. There are no science test scores for Indonesia, Iran, Kuwait, Morocco, Belgium. A list of participating countries with available data on instruction time in both subjects can be found in Appendix Table B2.1. Besides, I drop observations with more than one teacher in either science or math. Otherwise, there would be four or more observations per student. Since there is no information about why there is a change of teacher, I decide not to use these observations.

[5] In particular, I choose the first plausible value in both subjects since this is often used in the literature, for example in Rivkin and Schiman (2015). The results are robust to using the other four plausible values. This is in line with Jerrim et al. (2017) stating that using one plausible value or all five does not alter the results.

[6] Note that I reorder the four labels of the original variable in TIMSS such that the lowest category (1) equals "disagree a lot". The original variables used in TIMSS are shown in Appendix Table B2.2.

[7] Use M_k as the k^{th} of four variables, where μ_k is the mean and σ_k the standard deviation of each of the four variables. Then the normalized variables are $M_k^* = \frac{M_k - \mu_k}{\sigma_k}$. The final index is calculated by $M^* = \sum_k M_k^*/4$ which is then standardized.

The main independent variable of interest is *instruction time* either in math or in science, depending on whether the observation is for student *i* in math or science. The underlying question for teachers in order to determine this variable is the following: "In a typical week, how much time do you spend teaching mathematics to the students in this class? (minutes)" (TIMSS, 2017). The same question is asked for science. To make them comparable to other studies, I convert these variables into hours. Following Lavy (2015), I aggregate instruction time on the school-by-subject level. This is done to overcome potential problems due to sorting and tracking since schools might sort or track students into classes based on subject-specific instruction time. This can potentially lead to bias: an upward bias, for example, might occur when students who excel in math sort themselves into schools with more instruction time in math (Bingley et al., 2018).

Differences in instruction time occur both across and within countries. According to the curriculum data provided by TIMSS for each participating country, in some countries, the curricula prescribe the percentage of math and science lessons as a proportion of the total instruction time, while in other countries there is no such official document. Overall, 69.2 (61.5) percent of the countries indicate that the curriculum prescribes a certain percentage of instruction time in math (science). In some countries, it varies by state or school. Other countries define a range of percentages that should be devoted to instruction time in a given subject.

The information about a student's background, obtained from the student questionnaire, includes the gender of the student (the variable *female*).

The data also contain information about the teacher: a variable *teacher female* (one if the teacher is female), the teacher's *age* as well as several indicators of the *teacher's qualifications*. As explained in section 2.2, one approach in the literature used to determine the quality of a teacher is the teacher value-added. However, teacher value-added cannot be measured with the TIMSS data since students and teachers are only observed at one point in time.[8] Therefore, I use teachers' formal qualifications: these can be seen as one of two parts of teacher quality (Nilsen et al., 2018).[9] The formal qualifications of teachers consist of their educational background, measured by the years of experience, the highest level of education,

[8] Besides, TIMSS only measures the cognitive abilities of students and not those of teachers. Thus, I cannot use teachers' skills or subject knowledge as a measure for teacher quality as Metzler and Woessmann (2012) or Bietenbeck et al. (2018).

[9] The competencies of teachers are the second part of teacher quality. TIMSS also assesses teachers' competencies in teaching by asking them about the collaboration with other teachers, their motivation, their satisfaction with their job, their level of preparation, and their confidence (Nilsen et al., 2018). For example, there is a number of item questions asking for the preparedness of the teacher in their subject. This measure covers a broad spectrum and teachers might be less likely to misreport the level of preparation if they can differentiate between several subtopics. However, this measure, like the other self-reported teacher competencies, might suffer from an endogeneity problem: a teacher might be better prepared if she has to teach several hours than if she only has to teach a few hours. This is why I only focus on the formal qualifications of the teachers.

i.e., highest degree, specialization in a subject, and participation in professional development (Nilsen et al., 2018).

Of the large amount of questions in the TIMSS teacher questionnaire, I use the following variables for teachers' formal qualifications in my analysis. Teachers are asked about the highest level of formal education they have completed (according to the ISCED classification[10]), whether they majored in math or science, and whether they specialized in math or science when their major or main area of study was teacher education. I generate the following variables: first, *major degree* indicates whether the teacher has a Bachelor's degree (or higher) and whether she majored in the relevant subject. Hence, this variable is an indicator of a teacher's knowledge about the subject: it can be assumed that a teacher knows the content of a subject when she majored in that subject. Studying mathematics as a major, for example, provides knowledge of the content required for teaching mathematics to students (Blömeke et al., 2016). Second, *education specialization* indicates whether a teacher has a specialization in the relevant subject if she has an educational background, i.e., a major in teacher education. This variable accounts for the fact that teacher education and pedagogy are also relevant. When teaching the fourth grade, it is especially important that teachers know how to teach and are good pedagogues (and not only know about the content of the subjects). The third aspect is whether a teacher has participated in *professional development* (PD) in the respective subject in the last two years. Categories of professional development are subject content, subject pedagogy/instruction, or subject curriculum. These teacher qualifications are subject-specific: for example, a teacher might have a Bachelor's degree (or higher) with a major in math, but not in science. Similarly, a teacher might have participated in professional development in science, but not in math. Hence, these qualifications can vary even within teachers. Fourth, the teachers are asked how many years they have been teaching. On average, teachers have been teaching for 17.4 years across all countries, with a maximum of 60 years. I generate an indicator variable for those teachers who have more than two years of experience (*high experience*) since previous literature has identified these as the relevant years (Rivkin et al., 2005; Clotfelter et al., 2007): teachers' performance with only one or two years of experience tends to be worse than the one of more experienced teachers. Since there could also be tracking into classrooms based on teacher quality dimensions, I aggregate these four teacher characteristics on the school-by-subject level based on the rationale given for instructional time.

Overall, 20 percent of the teachers have a Bachelor's degree (or higher) with the relevant subject as their main subject and 28 percent have an educational background with a specialization in the subject (see Table 2.1). About half of all teachers participated in professional development in the last two years. However, these numbers conceal differences between developed and developing countries. In developing countries, more teachers have a Bachelor's degree (or higher) with the relevant subject as their main subject (30 percent) and more

[10] The ISCED classification is suitable for indicating the level of education or the highest level of qualification. This facilitates international comparison.

teachers have an educational background with a specialization in the relevant subject (38 percent).

The data also contain information about the geographical area of the schools, i.e., whether it is a *remote* or urban region, and whether *tracking* is used. In particular, headmasters are asked the following question (separately for math and science): “As a general school policy, is student achievement used to assign fourth grade students to classes?”

2.4 Empirical Strategy

Identifying the causal effect of instructional time on student attainment requires exogenous variation which is unrelated to any unobservable differences in schools and students (Rivkin and Schiman, 2015). Therefore, I use a student fixed effects model and exploit within-student between-subject variation. Due to their panel-like nature, the TIMSS data are particularly suitable since they provide two observations per student: a student’s attainment as well as the instruction time are reported for math and science. Using a student fixed effects model controls for unobservable student characteristics, such as unobserved ability and school differences in both subjects (Rivkin and Schiman, 2015). The attractiveness of this model lies in the fact that the students taking two subjects have the same general skill level and that the school environment is the same for both subjects (Rivkin and Schiman, 2015). Hence, no heterogeneity in terms of ability, habits or school quality will introduce biases to the estimates. Only subject-specific factors might have confounding effects. Therefore, I include subject-specific effects in the specification.

Based on this method, my regression equation is presented in equation (2.1).

$$test\ score_{ijk} = \beta_1 H_{kj} + \beta_2 X_{ij} + \beta_3 Q_{kj} + \mu_i + \epsilon_j + \eta_k + u_{ijk}, \tag{2.1}$$

where *test score*$_{ijk}$ is the test score for student *i* in school *j* in subject *k* (*k* $\in$ math, science). H_{kj} is the instruction time (in hours) in school *j* in subject *k*. X_{ij} are student characteristics of student *i* in school *j* and Q_{kj} are teacher characteristics in school *j* in subject *k*.[11] μ_i are student fixed effects, capturing for example unobserved student and family background. η_k are unobserved subject-specific characteristics and ϵ_j are unobserved school characteristics. Controlling for student fixed effects already controls for school fixed effects.

The effect of instruction time might differ according to the quality of the teacher. An additional hour of instruction by an unqualified teacher or a teacher who does not know the subject matter well might not result in higher test scores since students might not learn more during this time. It might be more important how time at school is spent, how good teachers are at

[11] However, most student characteristics, such as gender or the number of books in the household, cannot be used as X_{ij} since these characteristics are the same across both subjects and hence do not change within students. Teacher characteristics, however, can change since a student might have different teachers in the two subjects or since teachers might have a qualification in one subject, but not in the other one.

teaching and how motivated students are to learn and not so much the absolute amount of time (OECD, 2014). To assess this, I interact the instruction time variable H_{kj} with a quality indicator Q_{kj} of the teacher, measured by her formal qualifications:

$$test\ score_{ijk} = \beta_1 H_{kj} + \beta_2 X_{ij} + \beta_3 Q_{kj} + \beta_4 H_{kj} Q_{kj} + \mu_i + \epsilon_j + \eta_k + u_{ijk}. \tag{2.2}$$

The formal teacher qualifications consist of participation in professional development in the relevant subject, a Bachelor's degree (or higher) with the relevant subject as a major subject, teacher education with a specialization in the subject, as well as teaching experience as described in section 2.3.2. Regressions are run separately for each qualification.

Although I can difference out many unobservable factors with this approach, some unobserved factors that may confound the estimates could remain: students might choose themselves which school they want to attend or are assigned to schools based on subject-specific time of instruction (Lavy, 2015), which would invalidate the identification strategy. An upward bias can occur if students with a high level of interest and ability in math sort themselves into schools that offer more instruction hours in math. A downward bias, on the other hand, can occur if students with low ability in math (or their parents) choose schools offering more math instruction hours because they need more instruction time to understand the content (Bingley et al., 2018). However, this is not an issue here, since instruction time is measured at the school-by-subject level. Thus, this approach helps to overcome the bias that might arise from the non-random allocation of instruction time. Tracking also appears to be more common within and not across schools, and often in higher grades when students choose a specialization (Lavy, 2015). Reassuringly, less than one fifth of the schools in the sample use tracking as a school policy.

Nevertheless, it must not go unnoticed that the estimates of the causal effect of instruction time on educational attainment might not be unbiased (Rivkin and Schiman, 2015) and that the strength of the identification of causal relationships might be lower than in RCTs, regression discontinuity designs, or instrumental variables. Besides, measurement error might be present in the self-reported measure of instruction time. A fixed effects model might reinforce these measurement errors (Angrist and Pischke, 2008). Teachers who teach only one subject might report more accurately the time they actually teach to their class than teachers who teach two subjects. From Appendix Table A2.2, one can see that there is a slight difference in instruction time between those students with one teacher in both subjects and those with two teachers. Due to measurement error in the explanatory variable, the estimated effects might hence be downward biased (Glewwe and Kremer, 2006; Metzler and Woessmann, 2012). Consequently, my results would underestimate the true effect of instruction time and can therefore be seen as a lower bound effect.

A potential limitation of this strategy is that the effect of an additional hour of instruction is assumed to be the same in both subjects (math and science): β_1 does not vary by subject. Other studies, such as Cattaneo et al. (2017), Bingley et al. (2018) and Lavy (2015, 2020), also make this assumption.[12] Another assumption is that the impact of instruction time does not take into account spillovers from instruction time in other subjects: instruction time in science does not influence a student's test score in math, and vice versa. Positive spillovers could occur if more instruction time in science positively impacts students' math test scores.[13] However, such spillovers might not be a cause for major concern since the content of math and science lessons differs more in lower grades. At this stage, not many calculations and other mathematical concepts are used in science lessons. Nevertheless, the presence of positive spillovers, for example, would lead to an underestimation of the effect of instruction time.

2.5 Results

2.5.1 The Impact of Instruction Time

To determine the effect of instruction time on student achievement, one could regress test scores on instruction time using ordinary least squares (OLS). The results of this regression are presented in Appendix Table A2.3: an increase in instruction time by one hour *ceteris paribus* leads to a decrease in test scores of about 0.023 to 0.026 standard deviations. However, many unobserved confounding factors will most likely bias the results of such a regression, even after controlling for students' demographic characteristics (column 3). Hence, the negative effect of instruction time is probably due to the bias resulting from omitted variables.[14] If students sort (or are sorted) into schools or classes by ability, estimates might be biased upward (Bingley et al., 2018). On the other hand, compensatory teaching might lead to downward biased estimates (Bingley et al., 2018). Besides, instruction time might be correlated with unobserved factors influencing a student's achievement: parents can decide where the family lives and which schools their children should attend. This decision is likely to be based on the perceived quality of the school, which also includes instruction time (Bingley et al., 2018).

To overcome this problem, I perform a regression as presented in equation (2.1) in the previous section. The dependent variable is the test score of a student, either in math or science. Table 2.2, columns 1 and 2 present the estimates of the fixed effects model. Standard errors are clustered at the school level and both specifications include subject fixed effects. All regressions are weighted using senate weights.

[12] Cattaneo et al. (2017) and Lavy (2020) have established that this assumption is reasonable. Cattaneo et al. (2017) cannot reject the null hypothesis that the effects are the same, while Lavy (2020) argues that effects are similar in the subjects examined in his study (math, science and English) when comparing schools before and after a school finance reform in Israel.

[13] Other studies also assume no spillover effects: Bingley et al. (2018) and Lavy (2015, 2020) also assume that there are (almost) no spillovers: Lavy (2015, 2020) finds small spillovers from math to science, but overall no spillovers from one subject to another, while Rivkin and Schiman (2015) show that some spillovers exist in their model. Besides, Wu (2020) looks at spillovers but is also not able to examine spillovers from math to science and vice versa. He only finds "spillover effects from time spent in non-tested subjects" (p. 104) on tested subjects using TIMSS data.

[14] Lavy (2020) finds the same negative effect for a naïve OLS approach when analyzing instruction time in Israel.

The results show that an additional hour of instruction time increases students' test scores by 0.030 standard deviations. Hence, students benefit in terms of test scores from additional instruction time. This finding is in line with the results of previous studies. In column 2, teacher characteristics are added. The coefficient on instruction time remains the same and is thus robust to including these additional variables.[15]

By interacting instruction time with a student's gender, I examine whether there are heterogeneous effects with respect to a student's gender. The coefficient on instruction time for boys is slightly higher than in the baseline specification (Table 2.2, column 3), indicating that an additional hour of instruction time leads to an increase in test scores of male students by 0.037 standard deviations. The effect for girls is 0.024. Hence, boys seem to benefit more from additional instruction time. This finding is surprising since returns to schooling are often lower for boys than for girls, especially in low-income countries (Montenegro and Patrinos, 2014). A tentative explanation for this result could be that girls generally study more for school outside of school hours while boys study less in their time off from school. Consequently, boys might need to spend more time studying with a teacher than girls to improve their test scores. For girls, the time they spend on homework (and not instruction time) might play a greater role than for boys.

As explained above, additional instruction time can affect a student's attitude towards a subject. A student might become tired of a subject, leading to an aversion to the subject. Alternatively, a student might enjoy a subject even more when additional instruction time is used to deal with more specific content. Using the same specification as before, I therefore use the variable *like subject*, an index comprising four questions, as an alternative dependent variable. The results are shown in Table 2.2, column 4: they suggest that additional instruction time leads to a more positive attitude towards the subject.

2.5.2 Interaction between Instruction Time and Teacher Qualifications

The effect of instruction time might differ according to the quality of the teacher: if teachers do not actively use their time for teaching and if their teaching is not of high quality, the additional instruction time might not result in achievement gains for students. To assess this, I interact the instruction time variable H_{kj} with a quality indicator Q_{kj} of the teacher, measured by her formal qualifications, according to equation (2.2), specified in section 2.4.

The results in Table 2.3 suggest that having a teacher who participates in professional development (column 1), having a teacher with a teacher training background and a specialization in the subject (column 2) and having a teacher who completed the relevant subject as the main subject with a Bachelor's degree (or higher) (column 3) enhance the positive effect

[15] Student characteristics are not included in the fixed effects specification since these cannot be estimated because they are the same for each student in both subjects. Besides, including an interaction of subject with gender as well as mother's and father's education as further controls leaves the results unchanged.

of instruction time for students.[16] The coefficients on the interaction terms are statistically significant and range from 0.025 to 0.034 standard deviations. The impact of one hour more instruction time is 0.025 standard deviations if the teacher does not have an educational background with a specialization, while it is 0.050 when having a teacher with exactly such a background (column 2). Similarly, the effect is 0.058 standard deviations when having a teacher who has a Bachelor's degree (or higher) with the relevant subject as their major (Table 2.3, column 3).[17]

Figure 2.1 suggests that a teacher with a pedagogical background and specialization in the subject who teaches three hours has the same impact on student achievement as a teacher who teaches four hours but does not meet these criteria.[18] When a teacher teaches the same students for many hours, i.e., more than three hours, it is especially important for the effect of instruction time on test scores whether the teacher is highly qualified.

The results suggest that across all countries the impact of instruction time is enhanced by the fact that a teacher majored in the relevant subject, i.e., knowledge of the content, and that she has an educational background, i.e., pedagogy. However, in the case of having a teacher who has been teaching for many years, i.e., a more experienced teacher, the effect seems to be slightly reduced (Table 2.3, column 4). This is surprising in that a more experienced teacher is expected to know how to use the time in a way that benefits the students. Clotfelter et al. (2007), for example, find that more experienced teachers improve student achievement more effectively than less experienced teachers. However, the results do not change when the number of years of experience is restricted to 40 years to exclude outliers.

These results can complement the study by Rivkin and Schiman (2015). Instead of teacher qualifications, they examine the quality and environment of the classroom and find that this increases the effect of additional instruction time. Hence, it seems that both teacher quality and student behavior in the classroom play an important role.

In addition, I examine whether the interaction between instruction time and teacher qualifications also impacts a student's motivation towards the subject. The results are presented in Appendix Table A2.5, but do not offer statistically significant coefficients for all the teacher qualifications. The coefficients are even close to zero for teacher training background with a specialization, having a Bachelor's degree (or higher), and participation in professional development. However, those teachers with more than two years of teaching experience seem to lead to higher student motivation compared to those with less than two years of experience.

[16] Appendix Table A2.4 reports the R^2 values to show the variation explained in the teacher qualification measures after controlling for student and subject fixed effects. The table also reports the $1 - R^2$ to indicate the variation left after including student and subject fixed effects.

[17] The results hold also when including all qualifications at once: the coefficients on the teacher qualifications are slightly smaller, but still show the same sign and significance.

[18] The difference in the coefficients at three hours is statistically significant at the 10 percent level. The figure looks almost identical for having a teacher who has a Bachelor's degree (or higher) with the relevant subject as their main subject.

Although not statistically significant, this is what we would have assumed: teachers with more experience know better how to motivate their students and know how to make students like the subject.

2.5.3 Country Analysis

In addition, I examine whether the effect varies across country groups. Various countries from different continents participate in TIMSS. Hence, these countries differ in their culture, particularly in their educational culture and educational system. One main difference is that some participating countries are developing countries (such as Chile, Oman, and Saudi Arabia), while others are developed countries (such as France, the United States of America and Japan).[19] The results for developing and developed countries are presented in Appendix Table A2.6: the effect of instruction time on students' test scores is statistically significant and higher in developed countries (0.061 standard deviations, column 1) than in developing countries (not statistically significant, column 3). The magnitude of the coefficient on instruction time in developed countries is similar to the coefficient determined by Lavy (2015) for OECD countries.

In both country groups, the effect is smaller for girls, and even negative for girls in developing countries (Appendix Table A2.6, columns 2 and 4). A potential explanation for this is the fact that education for girls is still not taken as given in many developing countries. Hence, girls might react negatively to spending more time in school, knowing that they are needed at home for work or that they have to earn income that their families depend on (Glewwe and Kremer, 2006). Another reason might be that teachers spend the additional instruction time on boys and less on girls, leading to higher gains for boys than for girls.

More importantly, I also analyze how teacher qualifications interact with instruction time in these two groups of countries since the influence of teacher quality might vary from one educational system to another (Blömeke et al., 2016). Table 2.4 shows that in developing countries, having a teacher who participated in professional development (column 1) or having a teacher who completed the relevant subject as a main subject with a Bachelor's degree (or higher) (column 3) enhances the impact of instruction time.[20] The impact of instruction time is negative when having a teacher who does not have a degree or who did not participate in professional development, but it becomes positive when having a better qualified teacher: instruction time by a highly qualified teacher increases test scores by 0.016 (participation in professional development) or 0.027 (a Bachelor's degree (or higher) with the relevant subject as a major subject) standard deviations, while it seems to decrease test scores when having an unqualified teacher. One potential reason for this might be that in developing countries students are needed at home to work. If these students have to stay at school longer with a

[19] I group countries according to the WESP classification (United Nations, 2014) which uses an exchange rate based method for aggregation. I combine countries in transition and developing countries. In the following, the term "developing countries" also includes countries in transition.

[20] The coefficients on the interaction terms are larger in magnitude than in the specification including all countries.

teacher without good qualifications, they become distracted and unfocused, which leads to worse outcomes. The coefficient on educational background with a specialization, however, is much smaller and only statistically significant at the 10 percent level in developing countries (column 2). Hence, the results suggest that having studied the relevant subject as major subject is more important than having an educational background. In view of the observation that teachers lack adequate knowledge and that the quality of school education in developing countries is often low (Glewwe and Kremer, 2006), having studied the subject in question as a major could thus be an indicator of more substantial content knowledge about the subject and hence plays a moderating role for the impact of instruction time.

In developed countries, by comparison, having a teacher with an educational background (Table 2.5, column 2) seems to enhance the impact, as does having a teacher who completed the relevant subject as a major with a Bachelor's degree (or higher) (column 3). The coefficients are almost of the same magnitude and hence suggest that both subject knowledge and knowledge about pedagogical elements can enhance the impact of instruction time in developed countries. The coefficient on the variable participation in professional development is positive but not statistically significant.[21]

2.5.4 Robustness Checks

As a first robustness check, to rule out bias from unobserved teacher characteristics, I apply a within-teacher specification: I exploit the fact that students in primary schools are often taught by the same teacher in both subjects which is true for about 75 percent of the sample. In a within-student within-teacher specification, I can account for unobservable teacher characteristics in addition. Such unobserved teacher traits are, for example, a teacher's motivation or her pedagogical skills. If these unobservable teacher characteristics are correlated with instruction time and student achievement, this might bias the coefficients. For example, teachers who spend more time teaching math might also be more motivated which would result in an upward bias of the estimates. Hence, in the same-teacher sample, I can account for any teacher traits that influence the student's performance in both subjects, math and science, in the same way. More explicitly, all teacher traits that are subject-invariant, such as general motivation or pedagogical skills, can be controlled for. Hence, in this within-student within-teacher specification, I exploit variation within students and within teachers. Both, the coefficient on instruction time as well as the coefficients on the interaction terms including the measure of teacher qualifications, are robust in this variant of the specification (Table 2.6). The coefficients on instruction time are slightly larger than the ones in the main specification.[22]

[21] The equality of the interaction coefficients between developed and developing countries can be rejected for professional development and major degree (both at one percent level).
[22] In contrast, restricting the sample to the group of separate teachers in the subjects, the results only hold for professional development and experience, but not for education specialization and major degree (results are available on request).

Further, I assess the robustness of my results by piecewise excluding each country individually from the main analysis to see if outliers drive the results, both for instruction time and teacher qualifications. Both the coefficients on instruction time and teacher qualifications remain stable across all 39 regressions. The impact of instruction time on student achievement ranges from 0.021 (excluding Cyprus) to 0.035 (excluding Chinese Taipei) (detailed results are available on request).

I also investigate whether the relationship between instruction time and test scores is non-linear. Similar to Rivkin and Schiman (2015), I examine whether the returns to additional instruction time are diminishing. For this purpose, I add a squared term of instruction time (in hours) to the baseline specification (Table 2.7, column 1). The coefficient on the squared term is slightly positive but not statistically significant. Therefore, this result does not support the hypothesis of diminishing returns. The coefficient on the level of instruction time is slightly smaller but is still robust to including the squared term.

A threat to identification and a source of selection bias might arise from tracking students into different classes within school (Lavy, 2015). To assess whether the results are robust to this threat, I present results for two subsamples, that are characterized by whether schools pursue this school policy of tracking. Overall, the incidence of using tracking as a school policy is low with 14.36 percent across all schools in the sample. Columns 2 and 3 of Table 2.7 show that additional instruction time has a positive, statistically significant effect in both subsamples: reassuringly, the coefficient in the non-tracking sample is similar to that in the main specification (column 3).

Furthermore, I check whether there are differences with respect to the type of geographic area in which the school is located. In remote or rural areas, students or rather their parents often have no alternative in the choice of school as there is often only one school in places with 3,000 or fewer inhabitants. This mitigates the problem of non-random allocation to schools. The results in column 4 of Table 2.7 demonstrate that the effect in rural areas is about 0.037 standard deviations, which is slightly higher than the effect in the main specification.

The use of an alternative specification for instruction time and the teacher qualifications, i.e., the exact report of the teacher (no aggregation on school-by-subject level), yields a coefficient on instruction time that is almost identical to that in the main specification (Table 2.7, column 6). Similarly, the results also hold for the interactions with teacher qualifications (see Appendix Table A2.7). Overall, the findings in Table 2.7, columns 2 to 6, and Appendix Table A2.7 indicate that the results are robust to sorting and tracking.

Besides, the results for the impact of instruction time on a student's motivation and attitude towards the subject are robust when using the alternative index calculated according to Kling et al. (2007): the coefficients on instruction time are slightly larger than the ones from the index based on factor analysis (see Appendix Table A2.8).

A possible concern might be that some teachers do not teach the subject they are supposed to teach. For example, a teacher might be teaching science when she should be teaching math, possibly because she prefers science over math or because she has majored in science but is required to also teach math. To alleviate this concern, I only focus on the subsample of students who have two separate teachers for math and science. It can be assumed that in this case a teacher only teaches the subject that she is supposed to teach. Conversely, a teacher could spend more time teaching math if she teaches both subjects to the same students. The coefficient on instruction time in the subsample which includes only the students taught by two separate teachers remains almost unchanged (Table 2.7, column 7). This suggests that this does not bias the main effect that I estimate in Table 2.2.

2.6 Conclusion

Using a fixed effects model and within-student between-subject variation, I show that instruction time positively affects students' test scores. On average across all countries, I find that an additional hour of instruction time leads to 0.030 standard deviations higher test scores. More importantly, I find that teacher quality, measured by teachers' formal qualifications, such as teacher training with a specialization in the relevant subject, a Bachelor's degree (or higher) with the relevant subject as their main subject, and participation in professional development, plays a moderating role for the effect of instruction time on student achievement: the effect is larger for students with better qualified teachers. This is especially relevant in developing countries, where the effect of instruction time on student achievement is on average not statistically significant and close to zero. However, instruction time with a highly qualified teacher also increases test scores by 0.016 to 0.027 standard deviations in developing countries.

The estimates on instruction time are about the same magnitude as those of Rivkin and Schiman (2015), but slightly smaller than those of Lavy (2015) who finds effect sizes of 0.06 to 0.08 standard deviations. In line with Lavy (2015), I also find that the impact of instruction time is lower for developing countries, even close to zero and statistically insignificant.

Some further points are important to consider when interpreting my results. The first is to consider whether extending the instruction time in a subject increases the overall time that students spend in school and whether this is at the expense of reducing breaks and vacation time (Jarrett et al., 1998; Farbman, 2015). For example, more instruction time in math at the expense of instruction time in another subject, e.g., arts and music, might improve test scores in math, especially if the lessons are given by a highly qualified teacher. On the other hand, this could affect students' development in terms of creativity, physical activity and health, particularly in primary school, and especially students from lower socio-economic backgrounds since they often do not have access to voluntary education outside of school.

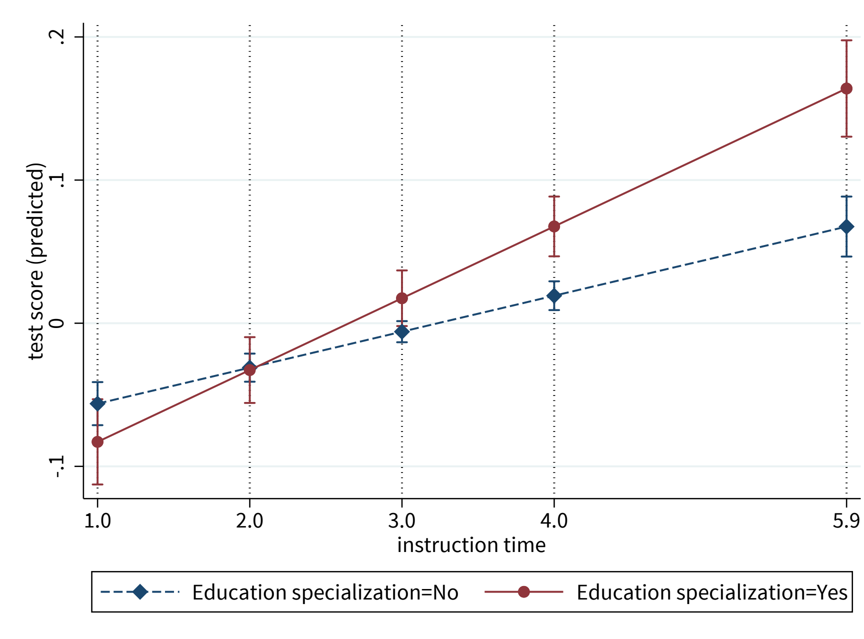
In addition to the education and training of a teacher, other school inputs and the behavior of teachers and students might affect the productivity of instruction time and make it rather endogenous (Lavy, 2015). Another factor might be how autonomously a school principal can make decisions. According to Lavy (2015), greater autonomy in decisions about hiring or firing teachers can lead to a better fit between schools and teachers, and thus to teachers making more effort. Besides, Barrios-Fernández and Bovini (2021) examine the intersection of school inputs in terms of instruction time and school institutions and governance by focusing on a reform in Chile which increased daily instruction, finding greater benefits due to longer instruction time for students in no-fee charter schools. As already mentioned, Rivkin and Schiman (2015) look at classroom quality and environment and find that this is important for student achievement. Their finding is in line with mine that higher qualified teachers have a larger impact on student achievement. Thus, the classroom environment in terms of student behavior as well as the quality of the teacher matter for the effect of instruction time.

Overall, my findings also relate to the discussion about whether money matters for schools (e.g., Card and Krueger, 1992; Hanushek, 2006; Woessmann, 2006; Jackson et al., 2021). The older literature often finds no clear effects that resources matter for achievement (Hanushek, 2006; Woessmann, 2006; Holmlund et al., 2010), but stronger effects depending on how the money was spent. Hanushek (2020), for example, states that it is more important to consider "how resources are used" compared to "how much is used" (p. 168). More recently, Jackson (2018) focuses on quasi-experimental studies that identify causal effects and finds that, on average, money matters. He further states that it is important to learn which types of increases in school spending are most important for student achievement. Hence, in the more recent literature, agreement on the fact that money matters for schools is growing (Jackson et al., 2021). My results contribute to this literature that it is especially beneficial for students in terms of achievement to increase instruction time, especially when taught by a high-qualified teacher.

When assessing the effectiveness of schools, more instruction time is not the only relevant component. However, my results and those of previous research suggest that instruction time is one of the key factors in promoting student achievement and that the quality of teachers, in particular the qualifications of teachers, can enhance the influence of instruction time on student achievement. The positive effect of instruction time on students' test scores and its interaction with teachers' qualifications is of particular importance for policy decisions. A slight increase in instruction time is most likely simple to implement. Hence, increasing instruction time would be an easy way to improve student achievement. However, since additional instruction time has to be financed, policymakers need to know whether this money is being invested effectively. My results suggest that it is the combination between instruction time and the quality of a teacher that is relevant to student achievement.

Figures and Tables

Figure 2.1: Marginal Effects Using *Education Specialization* as the Teacher Qualification Measure



Notes: Marginal effects using *education specialization* as the teacher qualification measure in a regression as in equation (2.2) with senate weights.

Table 2.1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Instruction time	2.959	1.538	0.017	10	216,716
Math instruction time	4.006	1.380	0.050	10	108,358
Science instruction time	1.913	0.913	0.017	9	108,358
Female	0.489	0.500	0	1	216,716
Teacher female	0.832	0.374	0	1	216,716
Teacher age	42.480	9.690	25	60	216,716
PD	0.478	0.479	0	1	216,716
Education specialization	0.276	0.447	0	1	216,716
Major degree	0.201	0.382	0	1	216,716
Experience	0.937	0.243	0	1	216,716
Tracking	0.144	0.351	0	1	185,184
Remote	0.324	0.468	0	1	210,176
Developed	0.627	0.484	0	1	216,716
Test score and *like subject* are standardized (mean 0, std. dev. 1).					

Notes: 4th grade sample in TIMSS 2015. Senate weights are used. PD stands for professional development.

Table 2.2: Baseline Results

Variables	(1)	(2)	(3)	(4)	(5)
	Test score	Test score	Test score	Like subject	Like subject
Instruction time	0.030***	0.030***	0.037***	0.046***	0.059***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)
Female x instruction time			-0.013***		-0.025***
			(0.002)		(0.004)
Observations	216,716	216,716	216,716	201,582	201,582
R-squared	0.923	0.923	0.923	0.607	0.608
Student FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	Yes	Yes	Yes	Yes

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015 in column (1) to (3); like subject index in math and science (from factor analysis) in 4th grade in 2015 in column (4) and (5). Instruction time is aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.1) and senate weights are used. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.3: Results for Teachers’ Formal Qualifications

Variables	(1)	(2)	(3)	(4)
	Test score	Test score	Test score	Test score
Instruction time	0.012***	0.025***	0.024***	0.034***
	(0.004)	(0.003)	(0.003)	(0.007)
PD x instruction time	0.030***			
	(0.004)			
Education specialization x instruction time		0.025***		
		(0.005)		
Major degree x instruction time			0.034***	
			(0.005)	
Experience x instruction time				-0.004
				(0.007)
Observations	216,716	216,716	216,716	216,716
R-squared	0.923	0.923	0.923	0.923
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Effect for high qualification	0.042***	0.050***	0.058***	0.030***
	(0.003)	(0.005)	(0.005)	(0.003)

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time and teacher qualifications are aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.2) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.4: Results for Developing Countries: Teacher Qualifications

Variables	(1)	(2)	(3)	(4)
	Test score	Test score	Test score	Test score
Instruction time	-0.030***	-0.007	-0.017**	0.023
	(0.009)	(0.007)	(0.007)	(0.017)
PD x instruction time	0.047***			
	(0.008)			
Education specialization x instruction time		0.013*		
		(0.007)		
Major degree x instruction time			0.044***	
			(0.009)	
Experience x instruction time				-0.023
				(0.017)
Observations	82,306	82,306	82,306	82,306
R-squared	0.940	0.940	0.940	0.940
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Effect for high qualification	0.016**	0.007	0.027***	-0.000
	(0.006)	(0.007)	(0.008)	(0.006)

Notes: Sample restricted to developing countries: Armenia, Chile, Chinese Taipei, Georgia, Hong Kong SAR, Kazakhstan, Rep. of Korea, Oman, Qatar, Russian Federation, Saudi Arabia, Serbia, Singapore, United Arab Emirates and Turkey. Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time and teacher qualifications are aggregated on school-by-subject level. Regressions run as in equation (2.2) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.5: Results for Developed Countries: Teacher Qualifications

Variables	(1) Test score	(2) Test score	(3) Test score	(4) Test score
Instruction time	0.058***	0.059***	0.059***	0.052***
	(0.005)	(0.004)	(0.004)	(0.006)
PD x instruction time	0.005			
	(0.005)			
Education specialization x instruction time		0.022***		
		(0.007)		
Major degree x instruction time			0.019***	
			(0.007)	
Experience x instruction time				0.009
				(0.006)
Observations	134,410	134,410	134,410	134,410
R-squared	0.908	0.908	0.908	0.908
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Effect for high qualification	0.063***	0.081***	0.078***	0.062***
	(0.004)	(0.007)	(0.007)	(0.004)

Notes: Sample restricted to developed countries: Australia, Bulgaria, Canada, Cyprus, Czech Republic, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Lithuania, New Zealand, Northern Ireland, Norway, Portugal, Slovak Republic, Slovenia, Spain, Sweden and the United States. Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time and teacher qualifications are aggregated on school-by-subject level. Regressions run as in equation (2.2) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.6: Results: Within-Teacher Specification

Variables	(1)	(2)	(3)	(4)	(5)
	Test score	Test score	Test score	Test score	Test score
Instruction time	0.035***	0.028***	0.032***	0.029***	0.039***
	(0.003)	(0.005)	(0.003)	(0.003)	(0.006)
PD x instruction time		0.012***			
		(0.005)			
Education specialization x instruction time			0.019***		
			(0.005)		
Major degree x instruction time				0.033***	
				(0.006)	
Experience x instruction time					-0.005
					(0.006)
Observations	161,588	161,588	161,588	161,588	161,588
R-squared	0.922	0.922	0.922	0.922	0.922
Student FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	No	No	No	No
Effect for high qualification		0.040**	0.051***	0.063***	0.034***
		(0.004)	(0.006)	(0.006)	(0.003)

Notes: Sample restricted to students with the same teacher in both subjects. Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time and teacher qualifications are aggregated on school-by-subject level. Regressions run as in equation (2.2) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.7: Robustness Checks

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Tracking: Yes	Tracking: No	Remote: Yes	Remote: No		
	Test score	Test score	Test score	Test score	Test score	Test score	Test score
Instruction time	0.017*	0.039***	0.027***	0.037***	0.026***	0.028***	0.028***
	(0.009)	(0.008)	(0.004)	(0.006)	(0.004)	(0.003)	(0.007)
Squared instruction time	0.002						
	(0.001)						
Observations	216,716	22,710	162,474	54,470	155,706	216,716	55,128
R-squared	0.923	0.943	0.919	0.916	0.925	0.923	0.930
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level in columns 1 to 5 and 7, not in column 6. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.1) and senate weights are used. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix

Appendix A2.1 Tables

Table A2.1: Factor Analysis: Factor Loadings and Scoring Coefficients

Variable	Factor loadings	Scoring coefficients
Panel A: Factor analysis math		
Enjoy learning math	0.935	0.280
Learn interesting things in math	0.836	0.250
Like math	0.959	0.287
Math is favorite subject	0.920	0.276
Panel B: Factor analysis science		
Enjoy learning science	0.932	0.276
Learn interesting things in science	0.862	0.256
Like science	0.961	0.285
Science is favorite subject	0.917	0.272

Notes: Eigenvalue: 3.340 (math), 3.373 (science); Proportion of variance explained by factor: 0.835 (math), 0.843 (science). Given the categorical nature of the variables, I use polychoric correlations to conduct the factor analysis.

Table A2.2: Instruction Time in Math and Science by the Same vs. Different Teachers in the Subjects

Variable	Same teacher	Different teachers	Diff.	p-value
Instruction time (both subjects)	2.95	2.99	0.05	0.00
Math instruction time	4.01	3.97	0.04	0.00
Science instruction time	1.88	2.02	0.14	0.00

Notes: Table shows the mean of overall instruction time and instruction time in math and science separately for the sample split by whether the students are taught by the same teacher in both subjects or by two different teachers, i.e., where teachers only teach a single subject. Senate weights are used.

Table A2.3: OLS Regression

Variables	(1) Test score	(2) Test score	(3) Test score
Instruction time	-0.026***	-0.023***	-0.014*
	(0.009)	(0.009)	(0.007)
Observations	216,716	216,716	213,074
R-squared	0.001	0.003	0.114
Student FE	No	No	No
Subject FE	Yes	Yes	Yes
Teacher controls	No	Yes	Yes
Student controls	No	No	Yes

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age* and student controls are *female* and *books*, i.e., the number of books at home as a proxy for the socio-economic status. Simple OLS regressions are run and senate weights are used. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2.4: Percent of Variation in Teacher Qualification Measures (Un-)Explained by Fixed Effects

Qualification	R^2	$1 - R^2$
PD	0.713	0.287
Education specialization	0.777	0.223
Major degree	0.800	0.200
Experience	0.878	0.122

Notes: Teacher qualifications used as dependent variables. The independent variables are student and subject fixed effects. Senate weights are used. The table shows the R^2 and $1 - R^2$ from each regression.

Table A2.5: Results for Teachers' Formal Qualifications: *Like Subject*

Variables	(1) Like subject	(2) Like subject	(3) Like subject	(4) Like subject
Instruction time	0.046***	0.045***	0.046***	0.030*
	(0.007)	(0.006)	(0.006)	(0.018)
PD x instruction time	-0.001			
	(0.007)			
Education specialization x instruction time		0.006		
		(0.007)		
Major degree x instruction time			-0.000	
			(0.008)	
Experience x instruction time				0.017
				(0.018)
Observations	201,582	201,582	201,582	201,582
R-squared	0.607	0.607	0.607	0.607
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Effects for high qualification	0.045***	0.051***	0.046***	0.047***
	(0.006)	(0.008)	(0.009)	(0.006)

Notes: Dependent variable: TIMSS like subject index in math and science (from factor analysis) in 4th grade in 2015. Instruction time and teacher qualifications are aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.2) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2.6: Country Analysis

Variables	(1)	(2)	(3)	(4)
	Developed countries		Developing countries	
	Test score	Test score	Test score	Test score
Instruction time	0.061***	0.068***	0.001	0.007
	(0.004)	(0.004)	(0.006)	(0.007)
Female x instruction time		-0.013***		-0.014***
		(0.002)		(0.004)
Observations	134,410	134,410	82,306	82,306
R-squared	0.908	0.908	0.940	0.940
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Regressions run as in equation (2.1) and senate weights are used. Countries are grouped into developed and developing countries according to the WESP classification. Developed countries: Australia, Bulgaria, Canada, Cyprus, Czech Republic, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Lithuania, New Zealand, Northern Ireland, Norway, Portugal, Slovak Republic, Slovenia, Spain, Sweden and the United States. Developing countries: Armenia, Chile, Chinese Taipei, Georgia, Hong Kong SAR, Kazakhstan, Rep. of Korea, Oman, Qatar, Russian Federation, Saudi Arabia, Serbia, Singapore, United Arab Emirates and Turkey. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2.7: Robustness: Instruction Time and Teacher Qualifications Not Aggregated

Variables	(1) Test score	(2) Test score	(3) Test score	(4) Test score
Instruction time	0.013***	0.023***	0.022***	0.033***
	(0.004)	(0.003)	(0.003)	(0.006)
PD x instruction time	0.025***			
	(0.004)			
Education specialization x instruction time		0.024***		
		(0.005)		
Major degree x instruction time			0.031***	
			(0.005)	
Experience x instruction time				-0.005
				(0.006)
Observations	216,716	216,716	216,716	216,716
R-squared	0.923	0.923	0.923	0.923
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Effect for high qualification	0.038***	0.048***	0.053***	0.028***
	(0.003)	(0.005)	(0.005)	(0.003)

Notes: Dependent variable: TIMSS student test score in math and science in 4th grade in 2015. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.1) and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2.8: Alternative Baseline Specification: *Like Subject*

Variables	(1) Like subject	(2) Like subject
Instruction time	0.049***	0.062***
	(0.006)	(0.007)
Female x instruction time		-0.026***
		(0.004)
Observations	201,582	201,582
R-squared	0.609	0.609
Student FE	Yes	Yes
Subject FE	Yes	Yes
Teacher controls	Yes	Yes

Notes: Like subject index in math and science (according to Kling et al., 2007) in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation (2.1) and senate weights are used. Clustered standard errors at school level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix B2.1 Tables

Table B2.1: List of Participating Countries

Country no. (TIMSS)	Country name	WESP classification	OECD country
51	Armenia	In transition	non-OECD
36	Australia	Developed	OECD
100	Bulgaria	Developed	non-OECD
124	Canada	Developed	OECD
152	Chile	Developing	OECD
158	Chinese Taipei	Developing	non-OECD
196	Cyprus	Developed	non-OECD
203	Czech Republic	Developed	OECD
208	Denmark	Developed	OECD
926	England	Developed	OECD
246	Finland	Developed	OECD
250	France	Developed	OECD
268	Georgia	In transition	non-OECD
276	Germany	Developed	OECD
344	Hong Kong SAR	Developing	non-OECD
348	Hungary	Developed	non-OECD
372	Ireland	Developed	OECD
380	Italy	Developed	OECD
392	Japan	Developed	OECD
398	Kazakhstan	In transition	non-OECD
410	Korea, Rep. of	Developing	OECD
440	Lithuania	Developed	OECD
554	New Zealand	Developed	OECD
928	Northern Ireland	Developed	OECD
578	Norway	Developed	OECD
512	Oman	Developing	non-OECD
620	Portugal	Developed	OECD
634	Qatar	Developing	non-OECD
643	Russian Federation	In transition	non-OECD
682	Saudi Arabia	Developing	non-OECD
688	Serbia	In transition	non-OECD
702	Singapore	Developing	non-OECD
703	Slovak Republic	Developed	OECD
705	Slovenia	Developed	OECD
724	Spain	Developed	OECD
752	Sweden	Developed	OECD
784	United Arab Emirates	Developing	non-OECD
792	Turkey	Developing	non-OECD
840	United States	Developed	OECD

Notes: Countries are grouped into developed and developing countries according to the WESP classification.

Table B2.2: List of Variables for a Student's Attitude and Teacher Qualifications

TIMSS Variable name	Question text	Answer choices (adjusted labels)
Like subject		
ASBM01A	I enjoy learning mathematics	*Answer choices for each question:*
ASBM01D	I learn many interesting things in mathematics	- Agree a lot (4)
ASBM01E	I like mathematics	- Agree a little (3)
ASBM01I	Mathematics is one of my favorite subjects	- Disagree a little (2)
ASBS04A	I enjoy learning science	- Disagree a lot (1)
ASBS04D	I learn many interesting things in science	
ASBS04E	I like science	
ASBS04I	Science is one of my favorite subjects	
Professional development: In the past two years, have you participated in professional development in any of the following?		
ATBM09A/ATBS08A	Mathematics/science content	*Answer choices for each question:* Yes *or* No
ATBM09B/ATBS08B	Mathematics/science pedagogy / instruction	
ATBM09C/ATBS08C	Mathematics/science curriculum	
Specialization: During your <post-secondary> education, what was your major or main area(s) of study?		
ATBG05AC	Mathematics	*Answer choices for each question:* Yes *or* No
ATBG05AD	Science	
Specialization: If your major or main area of study was education, did you have a <specialization> in any of the following?		
ATBG05BA	Mathematics	*Answer choices for each question:* Yes *or* No
ATBG05BB	Science	
Experience		
ATBG01	By the end of this school year, how many years will you have been teaching altogether?	< years >
Degree		
ATBG04	What is the highest level of formal education you have completed?	- did not complete upper secondary education (ISCED level 3) - upper secondary education (ISCED level 3) - post-secondary, non-tertiary education (ISCED level 4) - short-cycle tertiary education (ISCED level 5) - Bachelor's or equivalent level (ISCED level 6) - Master's or equivalent level (ISCED level 7) - Doctor or equivalent level (ISCED level 8)

Notes: Questions on students' motivation and attitude (student questionnaire) as well as on teacher qualifications (teacher questionnaire) from 4th grade sample in TIMSS 2015.

3 Can Predicting Dropout in Social Programs Increase Program Returns?*

3.1 Introduction

Dropout rates from social programs often exceed 30 percent, posing a considerable threat to a program's cost-effectiveness (Heckman et al., 1999). In the United States, youth mentoring programs commonly experience dropout rates of 40 percent on average and can reach more than 80 percent for vulnerable groups.[1] This is also true for other countries like Germany: for instance, in the mentoring program *"Rock Your Life!"* (RYL), examined in this paper, the dropout rate stands at 35 percent. Furthermore, Kosse et al. (2020) find a dropout rate of approximately 45 percent in the German "Baloo & You" mentoring program. As these mentoring programs have been shown to be effective in previous research (Kosse et al., 2020; Resnjanskij et al., 2024), the substantial number of dropouts raises the question of how mentoring program agencies can address this potentially costly issue. Since premature closures of mentoring relationships in youth mentoring programs have been linked to negative consequences for participants, such as negative emotions, lower self-esteem, higher likelihood to skip school, and increased alcohol consumption (Grossman and Rhodes, 2002; Karcher, 2005; Herrera et al., 2007; DeWit et al., 2016)[2], it is crucial to understand factors that are important for dropout. In addition, since mentoring programs in the U.S. and Europe are usually large and often publicly funded, it is of particular interest to use these funds in the best possible and most efficient way.[3] Therefore, program agencies should prioritize the identification of students at risk of

* This chapter is co-authored with Sven Resnjanskij.

1 In the case of "Big Brothers Big Sisters", the largest mentoring program in the U.S., Grossman and Rhodes (2002) report a dropout rate of 40 percent after 12 months. Kupersmidt et al. (2017) summarize evidence from over 300 mentoring programs and report an average dropout rate of 38 percent, with dropout rates of 46 percent among students with poor grades in school, and even higher rates for participants with a criminal record (55 – 81 percent).

2 DeWit et al. (2016) mention negative emotions triggered by rejection and abandonment as potential mechanisms for negative program effects. Grossman and Rhodes (2002) use an instrumental variable strategy and find diminishing treatment effects for early dropouts, and even an increase in alcohol consumption by mentees in mentoring relationships terminated within the first six months (see also Grossman et al., 2012). Karcher (2005) finds that low attendance of mentors predicts a decline in mentees' self-esteem and other behavioral competences. In a randomized trial on the effect of the "Big Brothers Big Sisters" program, Herrera et al. (2007) find that mentees in mentoring relationships that last less than three months are more likely to skip school than in the control group. However, random assignment in treatment and control groups alone does not identify the (causal) interaction effects between treatment and dropout, because dropout is determined post-treatment and most likely to be affected by unobservables that also affect outcomes of interest.

3 Currently, mentoring programs treat more than 2.5 million youth in the U.S. each year. The largest mentoring program, "Big Brothers Big Sisters", has served over 200 million children over the last ten years. In Germany, RYL has served around 10,000 adolescents since its foundation.

dropping out, not only in mentoring programs but also in other (educational) programs aimed at improving students' and children's lives, and implement (cost-efficient) interventions to prevent dropout after enrollment.

In this paper, we first seek to predict dropout from a mentoring program and identify key factors that contribute to dropout. In the second part, we illustrate a program agency's optimal behavior for targeting additional interventions to prevent participants from dropping out. This is achieved by leveraging the dropout predictions within a model designed to describe the agency's decision-making process. In particular, we consider interventions of varying costs that aim to reduce participants' dropout risk and assess the expected profit of a program agency based on the associated intervention costs. The analysis focuses on dropout from a mentoring program targeting disadvantaged adolescents in Germany. We use mentee survey data from the evaluation of a mentoring program that was conducted to study its effects on students' school performance and their transition into the labor market (Resnjanskij et al., 2024).

This paper proceeds in the following way: first, using machine learning (ML) algorithms (the Least Absolute Shrinkage and Selection Operator (LASSO) and the Random Forest), we predict participants' dropout risk after the first year of the mentoring program and identify key factors that influence dropout rates. Second, we set up a model to describe the cost-benefit trade-off that a program agency faces when confronted with program dropout of participants: the program agency faces an optimization problem where the objective is to minimize the dropout rate within the mentoring program under a constrained budget. Third, we integrate the predictions obtained from the ML algorithms to the cost-benefit model to show how the program agency can optimally react to the dropout threat through targeted interventions. Thus, we provide an interpretable tool for program agencies, assisting them in decision-making regarding whom to target with an additional intervention depending on its costs. We hereby examine several scenarios and assume different magnitudes of the costs associated with the related interventions.

Our findings indicate that the most important factors influencing dropout rates are the student's family environment (including parental employment and their living situation), the student's performance in math, and personal characteristics such as self-efficacy, and whether they engage in extracurricular activities. It is important to note that we do not make any causal claims in this analysis but discuss implications for program agencies and policies. In our setting, the ML algorithms have modest performance according to performance measures, such as the area under the Receiver Operating Characteristic (ROC) curve (AUC) value of 0.6 to 0.7. However, our study demonstrates that program agencies can use these predictions as a valuable tool in their decision-making process to determine which individuals should be targeted with additional interventions.

Using the dropout predictions generated by ML algorithms as inputs in the model, we calculate the proportion of participants who should be optimally targeted with interventions and assess the program agency's expected profit based on different intervention costs. As our model shows that the optimal allocation of an intervention of the program agency relies on participants' individual dropout risk, the agency's expectations of this dropout risk and their precision impact the expected program profit. By examining the agency's overall cost-benefit ratio when equipped with the necessary data and empirical methods for deriving accurate estimates of the individual dropout risk, we can quantify the additional benefits of precise predictions in terms of an economically relevant unit: expected program profit. Additionally, our analysis reveals that utilizing (imperfect) dropout predictions allows to target and allocate resources and use them efficiently compared to a scenario in which an agency targets all participants with an intervention. Using predictions, an agency can ultimately achieve a higher (expected) program profit than in the latter scenario in which all participants are targeted. Lastly, we demonstrate that leveraging accurate and precise predictions enables a program agency to obtain a higher expected profit compared to the current status quo of no intervention.

Our study contributes to the existing literature on dropout from mentoring programs and the identification of factors that predict dropout. Kupersmidt et al. (2017) examine premature match closures in mentoring relationships for youth in the U.S., using mentee, mentor, and program characteristics as predictors in logistic regressions. They show that engagement in risky health behavior, illegal or criminal activities, and school functioning problems, among others, are strong predictors for premature relationship closure. Using a multivariate Cox proportional hazard model, Raposa et al. (2019) find that similarity between mentees and their mentors in terms of race and ethnicity is a strong predictor for a long match. Grossman and Rhodes (2002) find that older adolescents and adolescents who experienced emotional, sexual, and physical abuse tend to terminate their relationships early. On the mentor side, they find that mentors with lower incomes and married mentors between 26 and 30 tend to be in shorter relationships. However, to our knowledge, we are the first to apply ML techniques to predict dropout from mentoring programs to identify important predictors. Hence, we add to this literature by examining dropout in a German mentoring program that is targeted at disadvantaged adolescents using novel ML tools to identify factors that determine dropout.

We also contribute to the literature on ML in economics, focusing on prediction problems (Kleinberg et al., 2015), and to the combination of ML methods with economic context and theory. Many studies have applied ML methods to policy problems: these have focused on identifying poorly performing tax audits with ML to guide policy (Battaglini et al., 2023) or on detecting tax evasion (e.g., Ruan et al., 2019) and insurance fraud (Kirkos et al., 2007). Moreover, studies have used ML to predict the presence of corruption to support policy responses (Ash et al., 2022), to assign refugees to economically optimal locations (Bansak et al., 2018), to predict gun violence and victimization (Heller et al., 2023), and jail-or-release decisions (Kleinberg et al., 2018). In addition, Marečková and Pohlmeier (2019) explore whether non-cognitive skills can predict unemployment and how their estimates can be used to assign youth and young adults to workforce training programs and psychological interventions. They also propose a new economic criterion to tune the threshold parameter for economic efficiency.

Most closely related to our study is the literature on using ML to predict dropout in education or labor economics: many studies have used ML techniques to predict dropout from high school or university (e.g., Kotsiantis et al., 2004; Dekker et al., 2009; Berens et al., 2019; Kemper et al., 2020; Buchhorn and Wigger, 2021). Berens et al. (2019) predict students at risk of dropping out from German universities to develop an early attrition detection system by applying decision trees and neural networks. Similarly, Kemper et al. (2020) also predict dropout from a German university and find that decision trees yield better results than logistic regressions, with a classification accuracy of 83 percent after the first semester. In an extension, Buchhorn and Wigger (2021) find that neural networks can outperform decision trees at the expense of interpreting the results. Sansone (2019) combines ML methods and economic theory to analyze dropout from high school. He argues that schools can improve their early warning systems to predict students at risk of dropping out by exploiting available high-dimensional data and ML methods.[4] We add to this literature by applying ML tools to the mentoring setting to predict individual dropout risk. Furthermore, we contribute by providing an economic framework to identify the trade-offs faced by a program agency that is confronted with substantial dropout by its participants and express the value of precise dropout predictions in terms of the expected profit that the program agency can achieve when using these statistical tools.

Lastly, we contribute to the literature on education interventions that aim at reducing dropout rates. The majority of this literature has focused on reducing dropout from school or university. Himmler et al. (2019), for example, implement a soft commitment device by having students sign a nonbinding agreement and commit to stay on course for their graduation. They find that this increases the likelihood of students to participate in and pass exams. Sandner (2015) evaluates a mentoring program as an intervention to reduce exam failure and finds that, indeed, the program reduces failure rates. Heller et al. (2017) study an intervention called “Becoming a Man” and find that it reduces crime and dropout, and increases school engagement in terms of GPA, days present, and enrollment status at the end of the year. While these studies have experimentally analyzed dropout, we add to this literature by setting up a theoretical model to examine decisions related to dropout. We focus on the program agency’s perspective and show how it can decide whom to target with an intervention.

The remainder of the paper is structured as follows: section 3.2 describes the mentoring program and the data set used in this paper. Section 3.3 introduces the empirical strategy to predict dropout and discusses ML methods to derive empirical counterparts of the expected dropout risk. Section 3.4 discusses the performance of different prediction algorithms and

[4] Another example in the context of education is the study by Wyness et al. (2023) who predict pupil achievement using prior attainment data and ML methods. Their ML predictions improve on teacher predictions but still exhibit large inaccuracies. Another example in social policy is the study by Chandler et al. (2011) that predicts youths with the highest risk for violence to target interventions such as participation in an advocacy program.

identifies important factors for dropout. In section 3.5, we introduce a model to describe the agency's choice set when reacting to the threat of early program dropout of participants. In section 3.6, we show the number of participants a program agency should optimally target with further interventions as well as the expected profit an agency is able to obtain using the ML predictions. We conclude in section 3.7.

3.2 The Mentoring Program and Data Set

This section describes the mentoring program RYL (section 3.2.1) and the evaluation study that provides the data set (section 3.2.2) as well as the characteristics of the sample of disadvantaged adolescents from the mentoring program (section 3.2.3). We also discuss reasons for dropout from the mentoring program (section 3.2.4).

3.2.1 The Mentoring Program

The mentoring program RYL was founded by a group of university students in 2008. Adolescents in eighth and ninth grade from lowest-track secondary schools[5] are assigned to university students who act as a mentor for the upcoming one or two years. These university students are mentors on a voluntary basis. In each participating program site, a self-governing university society is responsible for recruiting mentors and mentees. Mentors are selected using screening devices such as certificates of good conduct and personal interviews. Once the mentors are selected, the umbrella organization of RYL offers training and counseling for mentors on how to manage the relationship with their mentees, as well as training on the organization of the university societies.

For mentees, there is no individual screening of potential participants. However, university-student officials from the society visit schools in rather disadvantaged neighborhoods to recruit participants. Additionally, teachers and principals can recommend students for whom they think the program is most valuable. The students then receive information material and a consent form that they themselves and their parents must sign. Prospective mentees meet their prospective mentors during a first Kick-Off training after which the one-to-one mentoring relationships are formed, usually based on mutual preferences.[6]

The main goal of the mentoring program is to help adolescents to successfully manage their transition from school to professional life. The focus of the program is to provide career guidance, establish visions for the adolescents' future work life, and foster the mentees' self-esteem and trust in their own abilities. However, there are no guidelines on how the meetings should be structured. That means each pair of mentor and mentee can decide on the content

[5] *Hauptschule* or equivalent in the German system where different types of schools cater for different academic levels.

[6] For more information on the mentoring program, the umbrella organization, and the independent university societies, see Resnjanskij et al. (2024).

and activities during their meetings. The program aims for at least bi-weekly meetings. In addition to joint activities such as going to the cinema or the zoo together, mentors might provide advice for their mentees during stressful situations at school or with their families, help with occupational orientation, and with job applications.

3.2.2 The Evaluation Study

To analyze dropout, we use data from an evaluation which assigned eighth and ninth graders who applied for the mentoring program RYL to a treatment and a control group (Resnjanskij et al., 2024).[7] The treatment group received an offer to participate in the mentoring program, while the control group did not. Most students were randomly assigned to treatment and control group. However, randomization was only feasible in cases of local oversubscription, but oversubscription was not attained at every participating mentoring site due to the inherent variability in the number of applicants at each location. In cases where oversubscription was not achieved, it was not feasible to randomly assign program participation, leading to the assignment of mentors for all students at that site. In this paper, we only use data from students who obtained an offer to participate in the mentoring program. In addition to the students who were randomly assigned to the treatment group, our sample in this paper thus also includes students who were assigned a mentor without being randomized into the treatment group. In total, 11 mentoring sites across 12 different German cities participated in the evaluation.[8] The evaluation comprised two cohorts. The baseline survey for the first cohort took place between October 2016 and May 2017, and one year later for the second cohort. In this baseline questionnaire, collected through pen-and-paper surveys in schools, participating students are asked about demographic, socio-economic, and family characteristics (e.g., age, gender, socio-economic status (SES), and migration background). Additionally, students answered questions about their school performance, behavior, and economic preferences, such as risk and trust, and future job- and career-related attitudes, such as whether they know what they want to do later in life, and whether school seems important to them.

A follow-up survey conducted approximately one year after the program started collects outcome data. This follow-up survey's field period ended in June 2019 for the second cohort. Like the baseline survey, respondents filled out the follow-up surveys in their schools to achieve a high participation rate. If a pupil was absent from school when conducting the survey, teachers distributed the questionnaire to the respective student and sent it back via mail once the student had completed it. If students changed schools, they were contacted by phone. In total, 85.4 percent of the respondents in our sample completed the follow-up survey at school, 9.8 percent completed the survey on a different day and sent their questionnaires

[7] The study's design ensured that the selection of participating adolescents was similar to the selection of participants that is obtained in the absence of the evaluation.

[8] The schools that participated in the evaluation were located in Aachen, Berlin, Bonn, Chemnitz, Cologne, Duisburg, Essen, Hamburg, Leipzig, Luebeck, Lueneburg, and Mannheim. Duisburg and Essen are jointly organized by a single society.

by mail, and 4.7 percent could be reached via phone. Overall, the one-year resurvey rate exceeded 92 percent, independent of dropout.

3.2.3 Data Set, Sample Characteristics, and Selection into the Program

Overall, 442 adolescents participated in the baseline survey, of which 274 were originally offered participation in the mentoring program. Our final sample in this study comprises 254 treated individuals, i.e., those who have been assigned to a mentor and who filled out the follow-up survey. Thus, 20 students did not complete the follow-up questionnaire, and hence, the attrition rate is low at 7.3 percent.

Our main variable of interest is a binary variable indicating whether a student dropped out of the mentoring program after one year. 34.6 percent of the respondents dropped out early of the program: they stated that the relationship with their mentor did not last until one year after program start.[9] Among them, 87.3 percent stated that they last met more than six months ago. This variable stems from the follow-up questionnaire, while all other variables that we use are taken from the survey at baseline, i.e., before the start of the mentoring program.

Table 3.1 provides descriptive statistics for selected variables and compares our sample to a representative sample of German ninth-grade pupils (National Education Panel Study (NEPS), 2020, henceforth NEPS sample; see also Blossfeld and Roßbach, 2011). In our sample, 42.5 percent are male, and the adolescents are, on average, 14 years old, which is the regular age for attending eighth grade. Thus, adolescents in our sample are almost one year younger on average compared to the representative NEPS sample which is because some schools start the mentoring program in grade eight instead of grade nine. However, only ninth graders are sampled in NEPS. Due to the design of the mentoring program, students in our sample attend the lowest-track secondary schools in Germany at baseline. In the representative NEPS sample of ninth graders in Germany, more than one-third of all adolescents attend the highest, academic-track secondary schools (*Gymnasium*). Compared to the NEPS sample, where 35.7 percent of adolescents have a migration background, 57.9 percent of adolescents in our sample have a migration background.

We measure adolescents' SES with the number of books at home. This measure is commonly used in the literature (Schütz et al., 2008) and provides a good proxy for the families' educational, economic, and social background. In our sample, the share of adolescents with at most 25 books at home is 44.9 percent which is more than double the respective share in the NEPS sample (17.6 percent). In contrast, only 26.4 percent of our adolescents have more than 100 books (compared to 60.2 percent in the NEPS sample) and are therefore underrepresented in our sample. On average, 74.8 (59.5) percent of students in our sample state that their father (mother) is working part- or full-time, while the respective share is 94.0 (80.2) percent in the NEPS sample. The differences in the two samples regarding the family

[9] The program is set up to last one year, with a potential extension to a second year.

background are all highly statistically significant. Overall, these numbers corroborate that we have a sample of disadvantaged adolescents, demonstrating that, on average, the mentoring agency successfully offers its program to disadvantaged adolescents.

In addition, 14.6 percent of adolescents in our sample exhibit math grades that are so low that they are at risk of repeating a grade or failing the subject, while the share of these pupils is only 7.5 percent on average in the representative NEPS sample without students from the highest track. The difference is statistically significant. Surprisingly, the average school performance in math of the adolescents in our sample does not differ from that in the representative sample of students from the lowest-track schools. Focusing on students' non-cognitive skills in terms of the Big Five personality traits, there are no significant differences between the samples except for neuroticism: students in our sample exhibit slightly higher values for neuroticism. Lastly, adolescents in our sample engage less often in extracurricular activities, such as playing an instrument in an orchestra, engaging in youth groups or a volunteer fire brigade, compared to students in the representative NEPS sample.

3.2.4 Reasons for Dropout

In the follow-up survey, we asked respondents who indicated that they do not meet with their mentor anymore about why their relationship ended. In this section, we discuss reasons that mentees and mentors mention for dropping out.

Besides the surveys among mentees, we also conducted a survey among mentors which was administered from September to December 2018. The purpose of this survey was to gain an exploratory insight into the mentors' characteristics, including details about themselves and their studies, as well as their evaluations of their mentoring relationship. However, we did not invest the same effort in reaching out to the mentors as we did in contacting the adolescents since this survey was rather aimed at providing supplementary information that we did not plan on using in primary analyses. Given that the mentor sample consists of only 114 university students, with 31 indicating a premature closure of their mentoring relationship, we refrain from using this data during the subsequent sections of this paper.

The question eliciting reasons for dropping out was presented in a multiple-choice format, consisting of 12 (14) items for mentees (mentors) along with an open category. As a result, the reasons provided are not mutually exclusive. An overview of the reasons elicited is shown in Table 3.2. The two reasons for the discontinuation of the mentoring relationship most often mentioned by mentees are that they did not have enough time to meet (48.9 percent) and that they did not feel like meeting anymore (40.9 percent), while those mentioned most often by the mentors are that the mentee has not contacted the mentor anymore (48.4 percent) and that they did not have enough time to meet (25.8 percent). Mentor-mentee mismatch in terms of not getting along very well is a reason for the break-up of the relationship in only approximately 10 percent of the relationships.[10]

In some cases, it could be rational for mentees to leave the program: in theory, mentees may rationally decide to leave the program early if they perceive it as less effective or beneficial than initially expected. However, only a few mentees mention having had other expectations of the program compared to what it was (13.6 percent). Additionally, the fact that the mentor could not help is mentioned less frequently by mentees (8.0 percent). Similarly, only few mentees indicate that they have achieved everything during their relationship that they planned to do (3.4 percent). Thus, these descriptions suggest that the mentoring relationship has rather faded out.

Taking together these numbers with the evidence from previous studies that suggest negative effects from premature closure of relationships on the one hand, and positive effects of program participation in general on the other hand, further interventions to reduce dropout seem desirable.[11]

3.2.5 Pre-Selection of Predictors and Pre-Processing

To study potential predictors of dropout, we use variables from the baseline questionnaire, i.e., before the start of the mentoring program and before mentees are assigned to their mentors. Our data set contains a large set of background variables on the mentees. Self-reported data on grades and school hours are included. Table 3.3 lists the baseline variables that consist of the following categories: socio-economic predictors, parental background, life circumstances, help from others, lack of orientation in career choice, applications, school and kindergarten, social life, preferences, non-cognitive skills, personality traits, tests, future outlook, and self-assessed school behavior.

Missing values, noisy data, and inconsistent and superfluous data lead to low-quality data and will result in low-quality performance of any prediction model, including ML methods (García et al., 2015). Thus, data pre-processing is essential before starting to fit a model. As a first step, we address the presence of missing values. Overall, for the vast majority of variables in our sample, the number of missing values is small (Appendix Table A3.1).[12] For the small fraction of missing values in categorical variables, we recode the missing values to the mode category, i.e., the most frequently occurring category. For continuous variables, we impute the missing

[10] Not getting along very well might not be a reason for dropping out that often since mentees and mentors are matched based on mutual preferences.
[11] Of course, one has to keep in mind that the benefit for participants from not dropping out (versus dropping out) is not identified from studies evaluating the programs since dropout is endogenous in that setting.
[12] Four variables exhibit a larger fraction of missing values because the first cohort includes data from two pilot studies in which not all questions from the main survey were included yet. Self-reported grades exhibit a larger share of missing values because students at one specific school did not obtain grades in the previous school year.

values with the mean of the respective variable. For our dependent variable, dropout, we mainly rely on self-reported information regarding whether the relationship is still active. For the few cases where the mentee information is missing, we use administrative information on participation status from the mentoring sites.

Before applying different prediction algorithms to our data, we convert the categorical variables with more than two categories into dummy variables without omitting the baseline category since all categories can be included when using methods that automatically select the relevant subset of predictors. We end up with 89 variables from the baseline survey as predictors.

3.3 Machine Learning Algorithms to Predict Dropout

Predicting participants' dropout risk can be described as a standard prediction problem. Hence, prediction models taken from the ML literature provide a natural choice for estimating predictions. Using ML methods instead of more traditional methods bears the advantage that one does not have to specify the relationship between input and output beforehand, especially in settings where it is not clear which variables are important for predicting the outcome. For example, ordinary least squares (OLS) which minimizes the sum of squared residuals, focuses on unbiasedness and is thus not optimal for prediction problems (Kleinberg et al., 2015). In the presence of highly correlated predictor variables, OLS might still be unbiased but may have a large variance which can lead to overfitting. Overfitting refers to a situation where a model performs well on training data but does not generalize well to unseen data and can therefore lead to low prediction quality (Hastie et al., 2015). In addition, OLS tends to be unstable when the number of covariates (*p*) is close to the number of observations (*n*) and interpretation might become difficult since it will likely assign a non-zero coefficient to all explanatory variables (Hastie et al., 2015).

To overcome these shortcomings and to estimate predictions of dropout behavior, we use two supervised ML methods[13]: the LASSO and the Random Forest. Since we predict a binary event consisting of the two classes "dropout" or "no dropout", we use the binary classification techniques of the respective methods.

3.3.1 Prediction Algorithms

LASSO. The LASSO algorithm is a linear regression regularization technique. We use regularization (shrinkage) to reduce the variance which may come at the cost of introducing some

[13] In supervised ML problems, predictors X_i and the outcome Y_i are observed (Athey and Imbens, 2019). The use of labeled data sets distinguishes this ML strategy from unsupervised learning. These data sets are intended to "train" algorithms to correctly classify data or forecast outcomes. In contrast, unsupervised learning examines and groups unlabeled data sets. These algorithms are referred to as "unsupervised" since they identify hidden patterns in the data and the goal is to partition the data into subsamples or clusters (Athey and Imbens, 2019).

bias. Nevertheless, this enables us to derive more accurate predictions (James et al., 2015) of dropout which is our main objective. In particular, we use the LASSO regularization: LASSO penalizes the sum of the coefficients' absolute values, also called $L1$-norm. The idea is to penalize excessive complexity in the models (Varian, 2014) by adding an additional penalty term to the usual OLS minimization problem. To obtain the LASSO coefficients, one minimizes the following loss function[14] (James et al., 2015):

$$\min_{\beta,\lambda} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}|\beta_j| = RSS + \lambda \sum_{j=1}^{p}|\beta_j|, \tag{3.1}$$

where *RSS* denotes the residual sum of squares, *n* the number of observations, and *p* the number of predictors. y_i denotes the binary outcome of interest, dropout from the mentoring program, and x_{ij} is the value for the i^{th} observation for the j^{th} variable (i = 1, 2, ..., n and j = 1, 2, ..., p) (James et al., 2015). The tuning parameter λ controls the strength of the $L1$-penalty: as λ increases, the variance decreases and the bias increases. We can easily notice that if λ equals zero, we obtain the usual OLS formula. Since the dependent variable is binary, we apply a logistic regression and hence maximize the following penalized version (Hastie et al., 2016):

$$\max_{\beta,\lambda}\{\sum_{i=1}^{n}[y_i(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}) - \log(1 + e^{\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}})] - \lambda \sum_{j=1}^{p}|\beta_j|\}. \tag{3.2}$$

As a result of this regularization, some coefficients are exactly shrunken to zero (James et al., 2015). Since LASSO shrinks some coefficients to zero, it performs variable selection (i.e., model selection) and hence leads to sparser and simpler models. Thus, LASSO decreases the model complexity by decreasing the number of predictors and is, therefore, less prone to overfitting. This makes model interpretation easier than with many predictors and solves the curse of dimensionality, even in cases with a large number of variables that exceeds the number of observations ($p > n$).

To sum up, LASSO is an algorithm for model selection which identifies the variables with high predictive power. One of the main strengths lies in its capacity to automatically select the most relevant variables, especially if the researcher has no priors. This characteristic renders the LASSO less susceptible to overfitting, consequently enhancing predictive accuracy. Since we have a relatively large number of *p* compared to the overall *n* in our sample, LASSO seems a good fit for our study.

One shortcoming of the LASSO, however, is that the model selection is only consistent under the irrepresentable condition. This motivates our use of the adaptive LASSO (Zou, 2006; Ahrens et al., 2019). The nice feature of the adaptive LASSO is that it is variable-selection

[14] The loss function in any ML algorithm measures during optimization and training how well an algorithm is performing or rather how far off the predictions are from the actual values. This is also referred to as the objective function.

consistent under weaker assumptions compared to the "normal" LASSO for fixed *p* (Ahrens et al., 2019). The adaptive LASSO has the oracle property which means that "it performs as well as if the true underlying model were given in advance" (Zou, 2006, p. 1418). In other words, an estimator with the oracle property has the ability to select the true relevant variables consistently. The only difference to the "normal" LASSO is that we modify the penalty term by assigning differing weights to the coefficients (Zou, 2006). As weights, we use the absolute value of the inverse of the $\hat{\beta}$s from a ridge regression[15]:

$$\hat{\boldsymbol{w}} = \frac{1}{|\hat{\boldsymbol{\beta}}|^{\gamma}}.$$

Post-LASSO. In this model, we use the variables selected by LASSO and use them as covariates in a logit model to predict the outcome, i.e., dropout. Since Post-LASSO only chooses a subset of variables to utilize as predictors in an OLS or logit model, it is simple to interpret. Each variable's significance is well known (Sansone, 2019).

Tree-Based Methods (Random Forest). As an alternative classification method, we use a tree-based method, the Random Forest, which consists of growing many classification trees. We introduce classification trees in Appendix A3.1. We use this alternative method since trees, and thus also Random Forests, have the advantage that the researcher does not have to impose virtually any initial restriction on the functional form or the interaction of the predicting variables. Trees can inherently capture any kind of nonlinearity necessary for deriving accurate predictions. However, decision trees can suffer from high variance. A method for overcoming this problem is the Random Forest, an ensemble learning method. The idea of the Random Forest is to create several training sets, set up a prediction model on each training set, and then average the predictions that result from these models. The procedure is the following: first, on a chosen bootstrap sample of observations, one grows a tree. At each node of the tree, only a random subset of predictors (*m* out of *p*) is considered as split candidates, among which only one is used at each split (James et al., 2015). This procedure decorrelates the trees, i.e., reduces the correlation or similarity between individual trees. This process is repeated to grow multiple trees. To make the final prediction, a majority vote among the predictions of the individual trees is used. By averaging predictions across multiple trees, the variance decreases (James et al., 2015). Having a large number of trees results in an improved prediction accuracy, but also leads to a loss of interpretability (James et al., 2015). A Random Forest is thus not as easy to interpret as single decision trees. However, it is possible to estimate which variables are most important for the predictions, i.e., which variables contribute the most to improving the prediction accuracy (Varian, 2014).

To sum up, the intuition behind Random Forests is that by aggregating the predictions of numerous individual trees, the ensemble may better generalize, capture more complex patterns in the data, and minimize the impact of outliers. As a result of the randomization included

[15] Ridge regression is similar to LASSO, except that it adds a different penalty term to the loss function. We give more details in Appendix A3.1.

during the feature selection and bootstrapping processes, overfitting is decreased, and the diversity of the trees is increased. Compared to LASSO, it can handle both linear and nonlinear relationships and it is more robust to outliers.

Logit Regression Model. We finally consider a simple logit regression model for comparison even though it tends to overfit the data, especially in high-dimensional settings with a large number of predictor variables.[16] These models are fit by the maximum likelihood method, so we maximize the log-likelihood (Hastie et al., 2016) which is essentially the same as in equation (3.2) but without the penalty term. With that, we estimate the coefficients and then compute the probabilities of dropout for each observation.

3.3.2 Implementation Details

Training and Test Data Set. As is standard in any ML application, we split our data into training and test sets. While we use the training data set to fit prediction models, we evaluate the performance of the model predictions in the test data set to avoid overfitting. We use a Leave-One-Out procedure to generate a test sample of sample size *n* as follows: we hold out one observation from the sample and fit the prediction models in the training sample with the remaining $n-1$ observations. We then compute the out-of-sample prediction for the observation that has been held out from the estimation procedure. Next, we hold out a different observation of the sample. In turn, we fit the model with the remaining $n-1$ observations and compute the out-of-sample prediction. We repeat this procedure until every observation in the sample has been selected once to be held out from the sample, and *n* single-observation test data sets have been created. This results in an aggregated test data set consisting of all *n* observations with *n* predictions, each of which has been estimated out-of-sample (without seeing the data point for which the fitted models computed the prediction). This allows us to use the largest possible training sample of $n-1$ observations while making it necessary to fit all prediction models *n* times with a resulting high, but in our case feasible, computational burden.

Tuning. To achieve the best performance of the ML algorithms, we tune the hyperparameters of our two models.[17] In the case of the LASSO, we use 10-fold stratified cross-validation to tune the hyperparameter λ.[18] To do this, a grid of λ values is used and we then select the

[16] Since OLS might lead to calculating probabilities larger than one and smaller than zero, we use a logistic regression model instead of an OLS model as we are interested in the dropout probabilities.

[17] Hyperparameters are parameters that are set before the learning process starts. Hyperparameters are provided or chosen using a search process, as opposed to the model's parameters, which are immediately learned from the training data.

[18] The idea of 10-fold cross-validation is similar to the Leave-One-Out procedure explained before. The main idea behind cross-validation is to randomly divide the observations into *k* folds of equal size (James et al., 2015). Then one of the folds is used as a validation set, while the other $k-1$ folds are used for model fitting. The "error" is then computed, and the procedure is repeated *k* times. Averaging the *k* test errors yields the overall error. When we apply cross-validation, we use stratified cross-validation, i.e., we ensure that the data is split such that all splits are representative in terms of the location site where participants are from.

tuning parameter value for which the cross-validation loss is smallest (James et al., 2015). The cross-validation loss is deviance (Friedman et al., 2010) which corresponds to *RSS* in linear models. We subsequently refit the model using all available observations in the training sample with the selected value of the tuning parameter λ and use this model to predict the probability of whether a participant drops out early of the program for the out-of-sample observation, i.e., the observation in the test sample.

We also tune the following hyperparameters of the Random Forest: the number of variables that are randomly sampled at each split (*mtry*), the number of trees grown (*ntree*), and the minimal node size (*min.node.size*) which controls the tree complexity. We tune the hyperparameters for each of the 254 random forests. We proceed in the same manner as for LASSO, i.e., refit the model with the chosen values of the tuning parameters in the training set and use this to predict the dropout risk for the out-of-sample observation.

3.3.3 Performance Measures

The statistics literature suggests several measures for assessing the out-of-sample prediction quality of estimation methods. In this section, we describe these measures. Since our dependent variable is binary, we mainly use measures suitable for classification problems. All classifiers estimate a predicted probability for program dropout.

The first measure that we report is the accuracy. The accuracy measures the proportion of correctly predicted instances (i.e., those who drop out and are predicted to do so, and those who do not drop out and are not predicted to do so) over the whole set of observations. In particular, we add the true positives (TP) and true negatives (TN) and divide by the total number of classifications to obtain the overall success rate (Witten et al., 2017):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+TN+FN+FP}}.$$

In our application, TPs are those who drop out of the program ($W_i = 0$) and are correctly predicted as dropouts (Table 3.4). TNs are adolescents who do not drop out of the program, i.e., those who still have a mentor following the first year, and are correctly predicted as non-dropouts. False positives (FP) are those who do not drop out but are predicted to be dropouts. Adolescents who drop out but are predicted to still have a mentor after one year are false negatives (FN).

In case of unbalanced data, accuracy might be misleading. Therefore, two alternative measures for the quality of algorithms are sensitivity and precision. On the one hand, sensitivity yields the share of correctly predicted dropouts among all true dropouts, that is, the number of TPs divided by the sum of TPs and FNs. In other words, sensitivity is defined as the proportion of correctly identified positives (Han et al., 2012). One should prioritize sensitivity when aiming at minimizing FNs, i.e., those who are predicted to stay but drop out. This is the relevant measure if it is of interest to correctly predict a high number of dropouts among all true dropouts which is identical to minimizing the number of dropouts that are predicted as non-dropouts. On the other hand, precision is the share of correctly predicted dropouts among all predicted dropouts, that is, the number of TPs divided by the sum of TPs and FPs.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Lastly, the ROC curve is another performance measure that demonstrates a classifier's performance without taking the class distribution or error costs into account (Witten et al., 2017) which is a drawback of the previous measures. The ROC curve plots the sensitivity, i.e., the true positive rate (TPR), on the vertical axis, and $1 - specificity$, i.e., the false positive rate (FPR), on the horizontal axis. Specificity is the true negative rate showing the proportion of correctly identified negatives (Han et al., 2012). Thus, the ROC curve reveals the trade-off between the TPR and FPR (Han et al., 2012). A graph depicting the ROC curve usually also shows a diagonal 45-degree line which represents random guessing. Consequently, the further away the ROC curve is from the diagonal line (towards the upper left corner), the more accurate the model (Han et al., 2012). A related measure which results from the ROC curve is the AUC (Han et al., 2012). One can interpret the AUC as the probability that a randomly chosen dropout observation is more likely to be classified as a dropout than a randomly chosen non-dropout observation.

3.4 Performance of Algorithms

In this section, we first examine the performance of the different algorithms based on established measures in the statistics literature explained in the previous section. Then, we focus on the variables that are important for predicting dropout.

3.4.1 Results

First, we analyze the algorithms' performances based on established measures in the statistics literature explained in the previous section. Table 3.5 compares the predictive out-of-sample performance of the selected algorithms (Random Forest, LASSO, Post-LASSO, and logit) in terms of accuracy, sensitivity, and precision.[19] The higher the values and the closer to one, the better the performance of an algorithm. The results provide several insights. First, the performance of all algorithms seems rather modest. Table 3.5 shows that all the algorithms achieve accuracy values ranging from 0.6 to 0.7, and sensitivity values ranging from 0.1 to

[19] Accuracy, sensitivity, and precision reported in Table 3.5 rely on the assumption about the probability cut-off used to classify observations in dropouts and non-dropouts: the traditional cut-off used for classifying observations as a dropout is 0.5. For graphical illustration, we plot the distribution of the predictions in Appendix Figure A3.1.

0.5. Second, the Post-LASSO algorithm seems to perform better than the Random Forest algorithm in all three performance measures. In addition, the logit model performs similarly to the Post-LASSO, but the disadvantage is that the model cannot be estimated unless some arbitrarily chosen variables are dropped from the estimation.[20] This suggests that in our setting, ML algorithms do not outperform more traditional algorithms to a large extent.

Another performance measure is the ROC curves, which illustrate the performance of the algorithms for different cut-off values. They show how the choice of the best algorithms depends on the cut-off used to classify predictions as dropout. However, ROC curves provide no clear guidance on how the choice of the cut-off should optimally be made to select a specific point on the ROC curves. Points on the curve that are closer to the upper left corner, i.e., close to a sensitivity of one and $1 - specificity$ of zero, are preferable but it is not clear which point on the curve is the best. According to Figure 3.1, Post-LASSO again seems to perform best since its line lies closer to the upper left corner compared to the other algorithms, which is in line with the results in Table 3.5. The AUC measure also reinforces this pattern with Post-LASSO yielding the highest value of 0.7, closely followed by LASSO. The use of alternative algorithms ("normal" LASSO, ridge, and neural networks) led to no further improvements in prediction accuracy or sensitivity (results are shown in Appendix Table A3.2).

Thus, taken together, the ML algorithms have rather modest performance. Post-LASSO seems to perform best among the ML algorithms that we have chosen. Additionally, the logit model does not perform much worse than the ML algorithms.

3.4.2 Important Variables for Dropout Prediction

In this section, we want to learn more about which variables and information about the students are important for dropout and how the dropout rate depends on these variables. Therefore, we investigate the most important predictors in the LASSO and Random Forest algorithms in this section. The results could then be used as an initial starting point to target efforts for collecting variables that are most important for predicting dropout. For example, the agency might be interested in designing a (short) questionnaire for participants at the beginning of the mentoring phase to get a sense of the dropout risk. Such a short questionnaire might help the program agency to estimate participants' dropout risks and to target a specific share of participants with an intervention to reduce dropout in case funding for these additional interventions is limited.

The aim of these predictions is to make recommendations for program agencies and policy rather than generating causal relationships between the variables and dropout. In addition, we can learn more about whether the participants who benefit the most from the mentor-

[20] Out of 89 variables in total, seven variables are dropped in the full logit model: migrant first generation, city identifier MA, mother employment (do not know), father employment (do not know), mother university education (do not know), father university education (do not know), math performance (do not know).

ing program, i.e., the most disadvantaged students (Resnjanskij et al., 2024), also have a particularly high risk of dropping out early of the program.

Before turning to the most important predictors obtained from the ML algorithms, we present differences in selected baseline characteristics between participants who stay in the program and those who drop out early. Table 3.6 shows that those who drop out of the program tend to demonstrate poorer performance in math, possess lower overall grades, and display lower scores in cognitive reflection and effort tests. In addition, those who drop out of the program tend to have a migration background more often and have a lower socio-economic background than those who do not drop out, measured by their books at home. These findings are in line with Kupersmidt et al. (2017) who find that racial and ethnic minorities, migrants, and those having academic problems are at risk of premature closure of their mentoring relationship. Moreover, as expected, students who drop out of the program less often engage in extracurricular activities and seem less open to experience. The difference between the two groups is not statistically significant in all other four categories of the Big Five personality traits. Similarly, there is no significant difference by gender related to dropout.

Important Predictors. As opposed to the LASSO that automatically selects important predictors, the results from Random Forest have limited interpretability. Hence, it is important to obtain an overall summary of the importance of each predictor. To do so, we calculate the variable importance. The general idea is to take each predictor variable and randomly permute its values in the data while keeping the other variables as they are. Then, we can measure the resulting loss in accuracy on the out-of-bag samples, i.e., the damage to the predictive model when not having the true values of a given variable (James et al., 2015). We then calculate the average difference between the original data's prediction error and the prediction error after permuting each variable over all trees. Finally, we normalize this difference by the standard deviation of the differences (Breiman et al., 2018). This procedure yields the variables most important in the model and most important for accurate predictions of the response.

We present the importance of the variables according to LASSO and Random Forest in Figure 3.2. Since we use each observation once in the test sample, we run the Random Forest 254 times. We obtain 254 importance measures for each variable in the data set and calculate the mean value for each predictor. Panel A of Figure 3.2 lists the 20 most important predictors based on the Random Forest algorithm, sorted from the largest to the smallest importance measure. Panel A of Figure 3.2 shows that the variable "extracurricular activities" is the most important variable in the model.

Similarly, we can also focus on the variables selected by the LASSO algorithm. As for the Random Forest, we run 254 LASSO models and hence obtain 254 selections of variables. Panel B of Figure 3.2 shows the 20 variables selected most often in one of the subsets.[21] The figure

[21] Appendix Figure A3.2 and Figure A3.3 show the whole set of selected variables and their variable importance.

shows that the first 18 variables are selected in each round, i.e., in 100 percent of the subsets. When comparing these 20 variables most often selected by LASSO to those with the highest importance scores from the Random Forest, we see that more than half of the variables are among the most important variables in both models (shaded in blue), emphasizing their importance across models. These twelve variables are the following: willingness to self-assess math performance, participation in extracurricular activities, self-efficacy, application for an apprenticeship, mother employment (job-seeking), biological father, father employment (part-time), cognitive reflection test, openness, German speaking, classmates expect effort at school, and live with stepfather. Dividing these variables into different categories yields the following categorization: family environment, such as parental employment and the living situation, the student's performance in math, and personal characteristics, such as self-efficacy, openness, and whether the students engage in extracurricular activities.[22]

Dependence between Input and Response Variables. With the knowledge of the most important variables in the model, we also examine how these variables determine the dropout rate. To do so, we focus on the six most important variables with the highest importance score in the Random Forest and selected in each round by LASSO at the same time: the mother's and father's employment situation (whether the mother is job-seeking, and whether the father works part-time), the student's engagement in extracurricular activities, a student's self-efficacy, the willingness to self-assess her own math performance, and the application for apprenticeships.[23] To illustrate how these six baseline variables determine dropout, we use partial dependence plots (PDPs). PDPs help to understand the dependence between predictor (input) and response variables. They offer insights by visualizing the conditional probability $P(W = 0|X = x)$ for different realizations of X. They illustrate the partial effect of a feature on the predicted outcome of an ML model (Friedman, 2001; Molnar, 2021). To create a PDP, we first choose a predictor variable of interest. Next, we hold all other features in the model fixed and average the predictions over all possible combinations of these other features. By varying the value of the selected predictor variable across its range of values, we can predict the outcomes using the ML model for each value. This process allows us to observe how the predicted outcomes change in response to different values of the selected predictor variable while holding other features constant.

We investigate how these six previously mentioned variables are associated with dropout behavior in Figure 3.3.[24] The x-axes of the plots represent the values of the selected features, while the y-axes represent the corresponding predicted outcome. For those who were not engaged in extracurricular activities (Panel A of Figure 3.3), and whose mother was seeking a job at the time of the program start (Panel B), the risk of dropping out prematurely is around

[22] We do not elicit participants' time (in)consistent behavior. A related concept is patience which we measure in the survey. However, this variable is not selected either by the Random Forest or the LASSO.
[23] Parental employment categories are working full- or part-time, job-seeking, not employed (e.g., housewife/-husband or pensioner), or that adolescents do not know the employment situation of their parents.
[24] Figure 3.3 shows results from Random Forest. The results for LASSO/Post-LASSO are similar and are shown in Appendix Figure A3.4.

four to five percentage points larger than for those who engaged in extracurricular activities and whose mother was not seeking a job, respectively. Those with a father working part-time (Panel C) also have a lower risk of dropping out than those whose father does not work part-time. Panel D presents the link between students' willingness to report their math performance and the dropout risk. Participants who do not know or are not willing to assess their math performance are more likely to drop out.[25] In addition, those who have already applied for an apprenticeship have a higher dropout risk than those who have not yet applied (Panel E). A student with a higher self-efficacy score on the five-point Likert scale (i.e., a student with a higher belief in his or her own abilities to succeed in difficult situations) has a lower associated predicted dropout risk. On the other hand, students with low self-efficacy scores are more likely to drop out (Panel F).

Overall, the evidence of all major determinants of dropout points towards more disadvantaged participants being at higher risk of dropping out. Resnjanskij et al. (2024) find the program to have the highest effects on disadvantaged adolescents, whose higher dropout risk suggests that they might not be aware of the program's expected gains when deciding to drop out. Therefore, we also examine the PDP for the variable measuring the socio-economic background of a student, books at home. Appendix Figure A3.5 shows that more books at home, as measured by a higher value on the six-point scale, are associated with a lower predicted dropout risk. Conversely, disadvantaged participants with a lower socio-economic background are more likely to drop out. A potential reason for not selecting "books at home" in LASSO could be the nonlinear relationship between books at home and the dropout risk. Additionally, if this variable is highly correlated with others, then this variable might not be selected.[26]

3.5 An Economic Analysis of Program Dropout

Participants who drop out create direct and indirect costs for the program agency. These arise from administrative costs and forgone benefits of others who could have participated in the program. This section introduces a simple framework to analyze the trade-offs confronting the program agency and the participant. The agency compares expected benefits and costs before intervening to prevent potential dropout. Later program dropouts by an enrolled

[25] Surprisingly, students who report performing well in math (better than others) are slightly more likely to drop out of the program. This could mean that these children do not need support from mentors and hence choose to drop out of the program. However, the fact that the mentor could not help is rarely cited by mentees as a reason for dropping out (see Table 3.2).
[26] This phenomenon is a common problem regarding the interpretation of the variable selection by LASSO: if some variables are highly correlated, they serve as substitutes, and some variables will be used in one partition, not in another. Nevertheless, using very different variables can still produce fairly similar predictions (Mullainathan and Spiess, 2017).

participant are assumed to reduce the benefits of the agency because the program still faces the direct or indirect costs related to the initial enrollment before dropout.[27]

Figure 3.4 presents the timeline of both agents' choices (the program agency and the (potential) participant) in the model and emphasizes at which stage predictions of dropout risk enter the choice problem of the program agency. The initial decision by an adolescent is whether to apply for program participation or not. This is not modeled in our framework, and we take this as exogenously given. Next, the program agency decides on an applicant's enrollment. In this study, we assume that a program agency enrolls all applicants since there is enough supply of mentors. After enrollment of the applicants, the program agency can decide whether to intervene further and reduce the dropout risk, conditional on enrollment. To avoid unnecessary costs and the misuse of resources that might arise since some participants do not drop out anyway, the program agency needs predictions of the participants' dropout risk. In the final step, the participant can choose to drop out early, whether the agency intervened or not. Section 3.5.1 introduces a simple Roy model to describe a participant's dropout choice. Section 3.5.2 introduces a model to analyze the implications of the choice faced by the program agency when confronted with the possibility of program dropout by participants: the program agency can engage in active measures to reduce the dropout rate of participants; for instance, by providing stronger supervision, additional incentives, or other behavioral interventions. A potential intervention strategy is to provide comprehensive information about the program, including clear guidelines and structure, along with emphasizing the potential benefits for participants. This intervention approach directly addresses the issue of incorrect expectations by participants and minimizes the psychological costs associated with disappointment. Additionally, by providing a realistic understanding of the time commitment required from adolescents, the participation costs for mentees could be reduced. Incentivizing participation through the provision of vouchers could enhance motivation among participants, thus increasing the value they derive from engaging in the program (consumption value). To address the issue of mentees or mentors ceasing contact, the program agency could implement reminder systems for both mentees and mentors. These reminders serve as prompts to maintain regular communication and can also assist mentees who exhibit time inconsistency in their behavior. By alleviating the burden of scheduling meetings, these reminders might also reduce the psychological costs for participants. Furthermore, offering full supervision of the mentoring relationship by the agency can address self-control challenges and further minimize participants' psychological burdens. All these measures are costly and need to be assessed against their expected benefits. Targeting them to at-risk participants avoids unnecessary costs from spending resources on participants who would not have dropped out anyway.

[27] Note that in this paper, we do not intend to model the initial selection of applicants. We assume the pool of potential program applicants to be exogenous for the program agency.

3.5.1 A Roy Model for Participants' Dropout Decision

Participant *i* has the choice to complete the program $(W_i = 1)$ or to drop out $(W_i = 0)$ and receives the corresponding outcomes $Y_{i,1}$ or $Y_{i,0}$. The participant's benefits, e.g., from improved labor-market prospects in case of successful program completion, are denoted by $B_i = Y_{i,1} - Y_{i,0}$. We use V_i to denote the (subjective) costs linked to completing a program. Following the intuition of a generalized Roy model, participants choose to drop out if the subjective costs exceed the benefits:

$$W_i = \begin{cases} 1 \text{ if } Y_{i,1} - Y_{i,0} \geq V_i \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

The utility from completing the program is denoted by $U_i = (Y_{i,1} - Y_{i,0}) - V_i$. In accordance with Eisenhauer et al. (2015), we assume that other, subjective benefits or an immediate consumption value of the program (unrelated to $Y_{i,W}$), are incorporated in the subjective cost term of participation (V_i). For instance, immediate consumption value derived from program participation enters the cost term negatively, whereas psychological costs of participation arising from the lack of self-control, impatience, or discomfort from being exposed to an unknown situation positively contribute to the costs.

The probability of a potentially negative utility, i.e., $Pr[(Y_{i,1} - Y_{i,0}) < V_i]$, determines the probability of observing a dropout $\pi_{i,0} \equiv E[W_i = 0]$. In the next section, we analyze the case in which the program agency is able to invest additional resources to reduce dropout by lowering the participation costs V_i for an enrolled individual while leaving the content of the program unchanged.

3.5.2 Cost-Benefit Considerations of the Program Agency Related to Dropout

Let $Z_i \in \{0, 1\}$ denote the program status of an applicant *i*, where $Z_i = 1$ represents the status of an enrolled applicant, and $Z_i = 0$ describes an applicant who never got an offer to enroll. After being enrolled in the program $(Z_i = 1)$, a participant drops out with probability $\pi_{i,0} = 1 - \pi_{i,1}$, where $\pi_{i,1}$ denotes the probability of completing the program $(E[W_i = 1])$. As the dropout by a participant is observed only after enrollment, the agency must form a priori expectations about the probability of dropout $\hat{\pi}_{i,0}$.

The benefits of the program agency derived from participant *i* depend on the participant's completion of the program with higher benefits for a participant who completes the program $(B_1 > B_0)$. We assume that the benefits for the program agency are constant across all participants: in the event of a participant successfully completing the program, the agency receives a benefit of B_1, while in the case of non-completion, the benefit is B_0.[28] These benefits may come from multiple sources such as the possibility to obtain additional funding, media attention, reputation, and the long-term success of the participant. We assume that an applicant who has never been offered program access and a participant who drops out, provide similar benefits for the program agency ($B_0 = B(W_i = 0|Z_i = 0) = B(W_i = 0|Z_i = 1)$). The program benefits derived from an enrolled participant are given by the following switching regression:

$$B_i^P(W_i) = (1 - W_i) \cdot B_0 + W_i \cdot B_1 = B_0 + W_i \cdot (B_1 - B_0).$$

Using $\pi_{i,0} = 1 - E[W_i = 1]$ and $\pi_{i,1} = E[W_i = 1]$ and *n* to denote the total number of initially enrolled program participants, the expected total benefits of the program read:

$$E[B^P] = \sum_{i=1}^{n} E[B_i^P] = \sum_{i=1}^{n} [\pi_{i,0} B_0 + (1 - \pi_{i,0}) B_1] = \sum_{i=1}^{n} [B_0 + (1 - \pi_{i,0})(B_1 - B_0)].$$

The program faces direct costs (C_z) for each participant enrolled in the program ($Z_i = 1$). These costs are, e.g., recruitment costs and the direct costs for enrolling the applicants. They are independent of the enrolled participants' later dropout choice.

In principle, the program agency has the option to restrict program access. By comparing the expected benefits $E[B^P]$ and the costs C_z, the program agency could decide whom to enroll and whom to decline.[29] However, restricting program enrollment might not be feasible. If there are no frictions on the supply side (i.e., mentors), social programs might be willing to enroll every applicant interested in joining the program. This is also due to ethical considerations since the program agency might not want to turn away applicants when there are available spots.[30] Moreover, restricting program enrollment to participants who have a low dropout risk might lead the agency to send away adolescents who would benefit most from participation (i.e., the most disadvantaged adolescents). In fact, in our application, 42.1 percent of low-SES mentees dropped out early from the mentoring program, while the share was only 28.6 percent for higher-SES mentees.[31] Thus, we focus our analysis

[28] We acknowledge the potential concern regarding the assumption of constant program benefits B_1 and B_0 across participants, particularly in light of the finding of heterogeneous effects by adolescents' SES by Resnjanskij et al. (2024). However, even when we relax this assumption to allow for differential benefits based on participants' SES – for example, assuming that low-SES adolescents derive greater benefits from program participation ($B_1 = 2$) than higher-SES adolescents ($B_1 = 1$) – the findings of this study remain unchanged. We show this in section 3.6.2.

[29] We present this maximization problem in Appendix A3.2.

[30] In fact, the mentoring program RYL promotes the development of one's full potential for pupils and young people.

[31] Measuring participants' SES using the same measure as in Resnjanskij et al. (2024) yields the same numbers as using less than 25 books at home as the low-SES indicator. In Resnjanskij et al. (2024), adolescents are classified as "low-SES" if they meet at least one of the following three criteria: (i) lack of educational support, indicated by the absence of a university-educated parent and a scarcity of books at home; (ii) lack of economic or time support, indicated by living with a single parent and having few books at home; (iii) lack of language or institutional support, indicated by a first-generation migrant background (i.e., being born abroad). Adolescents who do not meet any of these criteria are classified as "higher-SES".

on a program agency with access to interventions to prevent dropout of already enrolled participants instead of restricting initial enrollment.[32] As previously mentioned, the program agency could actively employ measures to reduce dropout, such as sending messages via phone to remind mentees and mentors to meet, giving out gift cards for joint activities, inviting mentors and mentees to social gatherings, etc. Since the program agency has the incentive to avoid unnecessary costs arising from investing in the reduction of dropout for program participants who would never drop out in the first place, a precise prediction of the individual dropout risk, $\hat{\pi}_{i,0}$, is central for cost-efficiently distributing these additional measures.

Intervention Choice with Full Mitigation of Dropout Risk. Different trade-offs arise depending on the measures available to the program agency. In this section, we assume that the agency does not restrict enrollment and leave the conditioning on $Z_i = 1$ implicit in the notation. In a simple version of the model, the agency has access to an intervention policy $t_i \in \{0, 1\}$ with costs C_t that reduces dropout to zero if applied $(t_i = 1)$. The costs associated with the initial enrollment (C_z) are sunk costs and do not alter the choice of the agency at that stage. The expected profit of the program as a function of the intervention is given by:

$$E[S^P(t_i)] = \sum_{i=1}^{n} E[B_i^P - C_z - t_i C_t] = \sum_{i=1}^{n} [B_0 + (1 - \pi_{i,0})(B_1 - B_0) - C_z - t_i C_t].$$

For every participant, the agency selects an intervention policy $t_i \in \{0, 1\}$ to maximize:

$$\max_{t_i \in \{0,1\}, \forall i} \sum_{i=1}^{n} [B_0 + (1 - \pi_{i,0} + t_i \pi_{i,0})(B_1 - B_0) - C_z - t_i C_t]. \tag{3.4}$$

The program agency will invest in the dropout intervention for participant *i* $(t_i^* = 1)$ if the expected benefits from the mitigation of the dropout risk $\pi_{i,0}$ to zero exceed the costs:

$$\underbrace{\pi_{i,0} \cdot (B_1 - B_0)}_{\text{Expected benefits of the intervention}} > \underbrace{C_t}_{\text{Costs}}. \tag{3.5}$$

As the dropout risk $\pi_{i,0}$ determines the expected profit, the correct prediction of the dropout risk is instrumental in deriving the optimal policy for the agency. Wrong predictions lead to suboptimal intervention choices and decrease the profit of the agency. Thus, how the agency

[32] We also set up a two-stage model, including the enrollment choice as a first stage (see Appendix A3.2). We decide not to use it in the analysis since the status quo is to enroll all applicants when there is enough supply (mentors). Based on supplementary calculations, the potential supply of mentors across Germany could amount to approx. 75,000 (using data on incoming students and the share of those who engage in voluntary activities such as youth and social work). A cohort of eighth graders in Germany consists of around 135,000 students. However, not all adolescents within the program agency's potential reach are interested in participating. The supply-side restriction may not be important when considering that adolescents participate voluntarily in the program, and many may decide not to do so. Thus, even if 44 percent of adolescents within the program agency's potential reach decide not to participate, there would still be enough mentors available for each mentee (see also Resnjanskij et al., 2024).

computes these expectations ($\hat{\pi}_{i,0}$), and their precision determines the program agency's profit.

Table 3.7 presents the program profits based on the model in equation (3.4). As before, we assume that the intervention completely prevents dropout, i.e., we assume that a participant's dropout risk is reduced to zero once the program agency intervenes. Table 3.7 shows that a program agency intervenes with further measures for adolescents that exhibit a high dropout risk, i.e., for those whose dropout risk is larger than the costs relative to the benefits $(C_t/(B_1 - B_0))$. In that case, the expected profit for the agency is $B_1 - C_t$, no matter whether the participant would have dropped out or not. In case of no intervention $(t_i^* = 0)$, the agency either obtains B_0 (if $W_i = 0$) or B_1 (if $W_i = 1$).

Intervention Choice with Partial Reduction of Dropout Risk. To relax the assumption that the intervention mitigates the dropout risk completely, we assume that the intervention t_i' decreases the dropout risk by a constant share δ, but does not eliminate it to zero. We can formalize the maximization problem as follows:

$$\max_{t_i' \in \{0,1\}, \forall i} \sum_{i=1}^{n} [B_0 + (1 - \pi_{i,0} + \delta t_i' \pi_{i,0})(B_1 - B_0) - C_z - t_i' C_{t'}], \tag{3.6}$$

with $0 < \delta \leq 1$. In that case, the program finds it optimal to intervene $(t_i'^* = 1)$ if:

$$\delta \pi_{i,0} \cdot (B_1 - B_0) > C_{t'}.$$

The expected program profits in Table 3.7 only change for $W_i = 0$ in case the agency intervenes $(t_i'^* = 1)$. For an intervention that reduces dropout by 50 percent (δ = 0.5), the expected program profit then amounts to $(B_1 + B_0)/2 - C_{t'}$ (see Appendix Table A3.3).

Implications of Dropout Risk for Optimal Program Behavior. Figure 3.5 illustrates how the dropout risk affects the program agency's optimal choice for participant *i* who has already been enrolled $(Z_i = 1)$. We assume that the dropout risk is completely mitigated if intervention t_i is implemented and partially mitigated by a factor δ if intervention t_i' is implemented. In Figure 3.5, S_{10}^P and $S_{10}'^P$ denote the program profit if neither t_i nor t_i' is implemented. The program's expected profit is $B_1 - C_z$ if the dropout risk is completely absent $(\pi_{i,0} = 0)$ and decreases in the dropout risk: the higher the dropout risk, the lower the expected profit. S_{11}^P demonstrates the situation if intervention t_i is implemented: the program agency's expected profit is $B_1 - C_z - C_t$. The program agency intervenes as long as $\pi_{i,0}$ is larger than $b = C_t/(B_1 - B_0)$. Figure 3.5 shows that the larger the enrollment costs C_z, the lower the y-intercept and thus the expected profit. Furthermore, the lower the intervention costs C_t, the lower the dropout risk at which the program agency should start intervening (intersection of S_{10}^P and S_{11}^P). Thus, more applicants receive the intervention.

Furthermore, Figure 3.5 illustrates the optimal choice if an intervention (t_i') only partially mitigates the dropout risk. The program agency will then intervene if the expected profit from implementing intervention t'_i, denoted by S'^P_{11}, is higher than S'^P_{10}. This is the case if the dropout risk exceeds point *a* on the horizontal axis in Figure 3.5. In contrast to intervention t_i, the program profit is still declining in $\pi_{i,0}$ if t'_i is implemented.

We could further investigate the scenario in which the program agency has access to both interventions and can choose among the interventions to maximize profits from the set: $(t_i, t'_i) \in \{(0,0); (1,0); (0,1)\}$. Figure 3.5 provides the intuition of the program's optimal behavior for the scenario in which $C_{t'} < C_t$.[33] The program agency will choose intervention t'_i to partially mitigate the dropout risk for individuals with a small $(\pi_{i,0} > a)$ to medium dropout risk $(\pi_{i,0} < b)$, whereas for high-risk participants the program agency will find it optimal to switch to the more expensive but also more effective policy t_i to fully mitigate the dropout risk.

Beyond the aforementioned intuition, the scenario with more than one available intervention does not add additional insights when combined with the empirically estimated dropout. Hence, in the following section, we limit our attention to the single-intervention policy scenarios as described in maximization problems (3.4) and (3.6) to keep our empirical analysis tractable.

3.6 Performance of Cost-Benefit Model

In this section, we present the results of the cost-benefit model, focusing on the number of participants targeted with dropout-reducing interventions and on the program agency's expected profit in five scenarios. We first present results according to our main assumption of a full reduction of the dropout risk (section 3.6.1). In section 3.6.2, we relax this assumption and assume that the intervention decreases the dropout risk but does not eliminate it to zero.

3.6.1 Main Results

In this section, we first present five different scenarios to illustrate the use of the dropout predictions and the model. Table 3.8 summarizes the decision rule employed by the program agency as well as the specific predictions that are used in each scenario.

In the first scenario, referred to as the "predictions" scenario, the program agency makes decisions based on the cost-benefit model using predictions from ML algorithms ("ML predictions") as well as predictions from the logit model ("logit predictions"). The agency acts

[33] If the costs of the less effective intervention t'_i are higher than the costs of intervention t_i that completely mitigates any dropout risk, the program agency will obviously always prefer the cheaper and more effective policy. Hence, we discuss the more relevant case with $C_{t'} < C_t$. Further note that the program agency will never implement both interventions simultaneously for a single participant, because implementation of t_i already mitigates any dropout risk perfectly. However, this does not rule out that the program agency will target different participants with different interventions depending on their specific dropout risk.

according to the decision rule $\pi_{i,0} > C_t/(B_1 - B_0)$ from equation (3.5), which implies to intervene for participants if their (expected) dropout risk is larger than the relative costs. In this scenario, the program agency uses (imperfect) predictions about participants' dropout probability and considers the costs of a potential intervention.

The second scenario, known as the "Oracle" scenario, assumes that the program agency has perfect knowledge about participants' actual dropout risk. This can also be interpreted as having perfect predictions about the dropout risk. Thus, we assume that $\hat{\pi}_{i,0} = \pi_{i,0} \in \{0, 1\}$ in this scenario and the program agency follows the decision rule from the cost-benefit model in equation (3.5).

The third scenario, called the "naïve" scenario, involves the program agency targeting all participants with a predicted dropout risk higher than 0.5, disregarding the costs of the intervention (intervene if $\hat{\pi}_{i,0} > 0.5$). The predictions $\hat{\pi}_{i,0}$ stem from the ML algorithms in this scenario.

Lastly, we examine two more trivial scenarios: the fourth scenario, referred to as the "intervention for all" scenario, assumes that the program agency targets all participants with an intervention without considering or knowing the expected dropout risk. Thus, the agency assumes $\hat{\pi}_{i,0} = 1$ and thus $t_i = 1$ for $\forall$ *i*. This might be the case if a program agency is willing to counter dropout through costly interventions although it does not know the expected dropout risk or how to obtain the predictions. In contrast, the fifth scenario, the "no intervention" scenario, implies that no participant receives an intervention. This is the status quo of the program. Hence, no predictions are used and $t_i = 0$ for $\forall$ *i*.

Figure 3.6 depicts the share of participants that a program agency can optimally target with an intervention at various cost levels across the five scenarios. The x-axis depicts the relative costs of the intervention, i.e., the costs of the intervention relative to the benefits: smaller values indicate inexpensive interventions, such as sending out reminders through automatic, standardized text messages (e.g., to remind mentees to meet with their mentor). In comparison, larger values indicate more expensive interventions, such as incentivizing students with vouchers for their participation or close supervision of the mentoring relationship. In this section, we assume that all interventions, independent of their respective costs, are successful in fully reducing the dropout risk to zero. The y-axis represents the share of participants targeted.[34]

First, in the "ML predictions" scenario, the program agency should target (almost) everyone, including those with a low dropout risk, when faced with low costs. For example, for relatively inexpensive interventions, (e.g., $C_t/(B_1 - B_0) =$ 0.2), the program agency should target 65.0

[34] In addition, Appendix Table A3.4 presents the absolute number of targeted participants for each scenario. In Figure 3.6, we show results for just one algorithm (LASSO) because results are quite similar – Post-LASSO seems to perform a bit better. Appendix Table A3.4 includes results for the other algorithms, and these do not change the interpretation.

percent of the participants (Figure 3.6, blue dashed line). However, as the costs increase, the number of targeted participants intuitively decreases according to the cost-benefit model. Once the program faces relative intervention costs ($C_t/(B_1 - B_0)$) of 0.4 to 0.5, the number of targeted mentees drops since relative costs are larger than the dropout risk, and only those at high risk of dropout should be targeted with an intervention (e.g., for relative costs of 0.8, the program agency should target only 6.3 percent of the participants). When using logit predictions (light blue dot-dashed line) instead of ML predictions, the share of participants targeted with an intervention declines more rapidly. Subsequently, the share of targeted participants remains almost stable over the cost distribution. It only slightly drops from 45.3 percent at relative costs of 0.1 to 30.3 percent at relative costs of 0.9.

Second, in the "Oracle" scenario where the program agency has perfect knowledge of who drops out and who stays, it would target the intervention to 34.6 percent of the participants, which corresponds to the share of individuals who would drop out early from the program (gray line). Thus, a program agency would target a higher share of participants in the "ML predictions" scenario compared to the "Oracle" scenario for low-cost interventions, while it would target a lower share of participants for high-cost interventions. The share of participants targeted according to the logit predictions, however, is much closer to the "Oracle" scenario than that according to the ML predictions.[35]

In the "naïve" scenario where all mentees with a dropout risk above 0.5 are targeted, about 24.0 percent of participants should receive an additional intervention, independent of its costs (light gray dotted line).

Lastly, as the names suggest, in the "intervention for all" scenario, the program agency targets 100.0 percent of the participants with an intervention (black dashed line). In contrast, in the "no intervention" scenario, zero percent is targeted with an intervention (dark gray dot-dashed line).

Next, we turn to the expected program profits. We calculate expected program profits for the cost-benefit model according to Table 3.7. For these calculations, we normalize the benefits B_1 to one and B_0 to zero. Appendix Table A3.5 shows the expected program profit for varying relative costs and Figure 3.7 illustrates a program agency's expected profits relative to the "no intervention" scenario which represents the status quo. We set the profits in the "no intervention" scenario to 100.0 percent across the full cost distribution (Figure 3.7, dark gray dot-dashed line). Comparing the "ML predictions" scenario (blue dashed line) to the "no intervention" scenario reveals that intervening for low- to medium-sized costs can lead to higher expected profits. For high-cost interventions, however, not intervening results in a slightly higher profit. Nonetheless, the difference in expected profit between these two

[35] Appendix Figure A3.1, Panel B also shows that about 50 percent of the logit predictions are close to zero and about 27 percent close to one. However, as the performance measures in Table 3.5 show, the logit model performs worse at correctly identifying dropouts or non-dropouts which also translates into lower expected profits as we will show in the following.

scenarios is minimal. It results from the algorithms' imperfect predictions as evident from comparing the "no intervention" scenario to the "Oracle" scenario (gray line). The "Oracle" scenario consistently yields higher expected profit compared to the "no intervention" scenario and the other scenarios. Even if the process of predicting dropout carries its own costs, the program agency can still justify these expenses by leveraging the higher expected profit obtained through targeted interventions. This increased profit can be used to cover the costs associated with employing an additional staff member or researcher dedicated to performing the calculations of the dropout risk. Thus, by investing in accurate predictions and strategic interventions, the program agency can optimize its outcomes.

The comparison between the "ML predictions" scenario and the "Oracle" scenario demonstrates that when the relative intervention costs are (close to) zero, the program agency can achieve a similar profit by targeting (almost) everyone with an intervention as if it had perfect knowledge. Even for high intervention costs, the program agency could still obtain profits that are almost as high as in the "Oracle" scenario: by using predictions from the ML algorithms, the program agency can secure at least 80 percent of the profit that it would obtain with perfect knowledge about the participants' dropout risks. This highlights the potential of using dropout predictions from ML algorithms in conjunction with the cost-benefit model for program agencies to make informed decisions about targeting interventions to prevent dropout, considering the associated costs. Furthermore, the results show that improving the quality and precision of predictions can enhance the agency's expected profit: getting close to the "Oracle" scenario represents the maximum achievable profit if the program had highly accurate and precise predictions of individual dropout. Thus, improving the quality of the predictions is pivotal and would benefit the program agency in terms of maximizing profit.

Furthermore, we compare the "ML predictions" scenario to the "intervention for all" scenario (black dashed line), where a program agency has no knowledge of a participant's dropout risk and targets all participants: the higher program profit observed in the "ML predictions" scenario as compared to the "intervention for all" scenario demonstrates the possibility of efficient resource utilization. Particularly for high-cost interventions, the "ML predictions" scenario yields higher expected program profit compared to targeting everyone. Intuitively, targeting interventions to all participants is especially costly if the intervention itself is expensive, resulting in less than 25 percent of the profit. Using (precise) predictions of dropout risk may also prevent a program agency from targeting mentees with a low dropout risk who may find additional interventions bothersome, such as receiving excessive reminders. Thus, these results show that even with imperfect predictions, the program agency is better off than without any predictions.

Comparing the "ML predictions" scenario to the "naïve" scenario (light gray dotted line), where all participants with a predicted dropout risk larger than 0.5 are targeted regardless of the intervention costs, shows that the "ML predictions" approach enables a higher expected profit across the cost distribution. This emphasizes the significance of considering the costs associated with interventions in the model.

Lastly, the expected profit calculated with predictions from the logit model is consistently smaller or equal to the one with ML predictions, indicating the benefits of using ML algorithms for dropout prediction. This can also be seen from the performance measures of the algorithms (Table 3.9): for smaller values of the costs (cut-offs), the LASSO model outperforms the logit model in terms of sensitivity, indicating its ability to correctly identify a higher proportion of TPs (dropouts). For larger values of the costs, the LASSO model exhibits better precision compared to the logit model indicating its ability to minimize FP cases. This is in line with the following observations: when the relative intervention costs are low, a program agency should focus on minimizing FNs which refer to cases where individuals are predicted to stay in the program but end up dropping out. In that case, the costs of failing to target individuals who eventually drop out (foregone benefit of $B_1 - C_t$) are larger than the intervention cost (C_t). Thus, it is of interest to predict more participants as potential dropouts and intervene for them, rather than risk failing to target individuals who eventually drop out. In contrast, when the costs are high, the program agency should aim at minimizing FPs, which refers to cases where individuals are predicted to drop out but actually do not. This approach is driven by the consideration that targeting too many participants with interventions can be financially burdensome. Given the higher costs associated with interventions, it is of interest for the program agency to predict fewer participants as potential dropouts, even if it means accepting a certain level of FNs.

Taken together, we show the value of predictions for the program agency in terms of expected profit and, more specifically, the value of precise predictions. Targeting interventions to at-risk participants, even with imperfect predictions, yields a higher expected profit for agencies compared to targeting all participants. In addition, we show that it is important to take the intervention costs into account instead of naïvely targeting participants with predicted dropout probabilities larger than 0.5.

3.6.2 Extensions

To relax the assumption that the additional intervention reduces dropout to zero, we could assume that the intervention decreases the dropout risk but does not eliminate it to zero. Thus, in what follows, we assume that an intervention halves the dropout risk (δ = 0.5 in equation (3.6)). We still assume a binary intervention, i.e., to intervene or not. In that case, the program agency cannot adjust the intensity of the intervention.

Figure 3.8 shows the share of participants that should optimally be targeted with an intervention according to the assumptions in the five scenarios. Compared to the analysis in the previous section, one can see that the share of targeted participants drops faster and reaches only a small fraction of about 6.3 percent at relative intervention costs of 0.4 and 0.0 percent at 0.5 relative costs ("ML predictions" scenario). The share of participants targeted with an intervention remains the same as before in the "naïve", the "intervention for all" and the "no intervention" since the decision to target participants with interventions does not depend on predictions and the costs in these scenarios. The "logit predictions" scenario exhibits a similar behavior to the previous analysis, albeit with slightly lower shares of targeted participants than before. The share of targeted participants also drops to zero at relative costs of 0.5 as in the "ML predictions" scenario and "Oracle" scenario.

Figure 3.9 shows the expected program profit under the assumption that interventions lead to a 50 percent reduction in dropout risk. For interventions with low relative costs, the expected program profit is reduced compared to the previous analysis. Despite this reduction, we can still observe that a program agency can achieve a higher expected profit by intervening compared to not intervening, even if the intervention only halves the dropout risk. This holds for both the "ML predictions" and "logit predictions" scenarios up to relative costs of 0.3. In the "Oracle" scenario, where perfect predictions are assumed, the program profit remains higher than in the "no intervention" scenario throughout the cost distribution until relative costs of 0.5, implying that intervening is a good choice. For expensive interventions, however, the agency's expected profit is the same for all scenarios except the "intervention for all" scenario. This is due to the fact that a program agency should optimally not target any participant anymore since costs are too expensive compared to the benefits of the intervention. Once again, targeting interventions to all participants appears unfavorable in terms of expected program profit for a program agency.[36] Therefore, the findings from the main model in section 3.6.1 remain valid up to relative intervention costs of 0.5, even if we assume a reduction of a fixed proportion.

Another concern might be the assumption of constant program benefits B_1 and B_0 across participants. However, relaxing the assumption to allow for differential benefits by participants' SES ($B_1 = 1$ for higher-SES mentees and $B_1 = 2$ for low-SES mentees) does not qualitatively change the results. As anticipated, the share of low-SES participants that is optimally targeted is higher in the "ML predictions" scenario as is the expected program profit (Appendix Figures A3.6 and Figure A3.7). In fact, the value of precise predictions becomes even more evident with this assumption in the "Oracle" scenario where the expected profit is larger than in previous results, relative to the "no intervention" scenario. Moreover, the program agency can still achieve a higher expected profit by making decisions according to the cost-benefit model compared to targeting all participants or compared to not intervening for anyone.

3.7 Conclusion

This paper provides an analysis of dropout of social programs from the perspective of a program agency. Using data from a mentoring program, it provides guidance for targeting interventions to reduce dropout in programs implemented to improve the labor-market prospects of disadvantaged adolescents, for example.

[36] Notice here that the "naïve" scenario yields the same expected profit as the "no intervention" scenario.

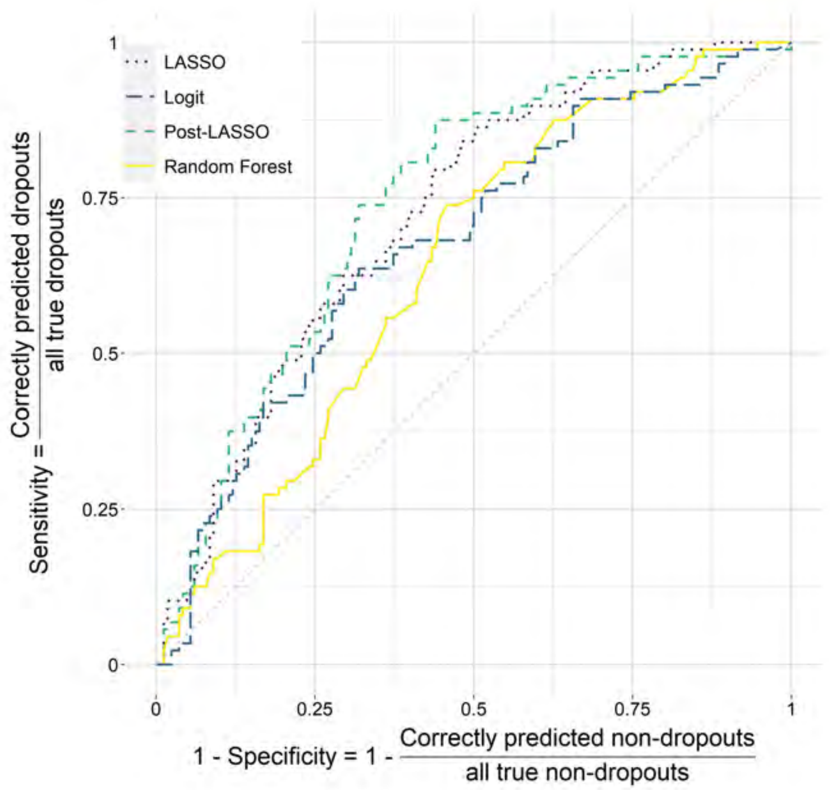
We analyze how a program agency can optimally respond to the threat of program dropout by intervening to reduce participants' dropout risk. First, we predict dropout from the mentoring program after the first year using different ML algorithms (LASSO and Random Forest) and identify important predictors of dropout. Overall, the performance of the algorithms in terms of accuracy and sensitivity is moderate. Second, we set up a model to describe the cost-benefit trade-off that a program agency faces when confronted with program dropout of participants and derive the optimal decision rule for a program agency to target additional interventions, depending on the size of the intervention costs and benefits ($\pi_{i,0} > C_t/(B_1 - B_0)$). We apply the predictions from the ML algorithms to the cost-benefit model to show how a program agency can use the predictions and react to the dropout threat. We examine several scenarios and show how many participants a program agency should optimally target with additional interventions depending on the associated intervention costs to maximize their expected profits. Moreover, we calculate the expected program profit that an agency could obtain and discuss how this depends on the precision of the predictions. Comparing different scenarios, we show that intervening to reduce dropout using predicted dropout risk is a valuable strategy for a program agency compared to no intervention, even if the predictions are imprecise. More precise predictions increase a program agency's expected profit. Thus, improving the quality of the predictions is pivotal and would benefit the program agency in terms of its profit. Furthermore, program participants could benefit in terms of avoiding negative psychological consequences and by improving their labor-market prospects.

Understanding dropout holds significance due to the typically large and often publicly financed nature of mentoring programs. Hence, it is of particular interest to optimize the use of these funds and implement (cost-efficient) interventions to prevent dropout after enrollment. This practice bears relevance not only for mentoring programs but also for various (educational) programs designed to enhance students' and children's lives, such as early childhood, tutoring, crime reduction, and health programs, as dropout prevention is a shared concern. Additionally, preventing dropout from any program is important to consider since total foregone benefits of a large national program may surpass those in controlled environments like randomized controlled trials. This could be attributed to an increasing share of dropouts that might emerge upon moving out of a controlled environment. Alternatively, it could be linked to an increase in the absolute number of dropouts resulting from a program scale-up, even if the share of dropouts remains constant. Hence, setting up strategies and interventions to reduce and prevent dropout is essential for any policymaker and researcher.

This paper uses modern prediction tools to derive empirical counterparts of the expected dropout risk, a crucial factor influencing the policy decisions of program agencies. The exact techniques from the statistical learning literature are not essential for the conclusions derived in this paper and more traditional methods may be used instead. However, using a logit model does not provide better predictions in this setting. Many program agencies, especially during the initial implementation phase, share the feature of having relatively few observations with many possible predictors. Precisely in these settings, ML techniques could provide the highest benefits over traditional methods as the researcher is more likely to find herself in a situation with a large set of predictors relative to a small sample size. Thus, the methods presented in this paper can serve as useful tools for program agencies to conduct pre-tests or pre-analyses helping them identify important variables for predicting dropout and determine the optimal number of participants to target with additional interventions. Hence, ML can be seen as a technique to complement more traditional economic models and estimations rather than a substitute.

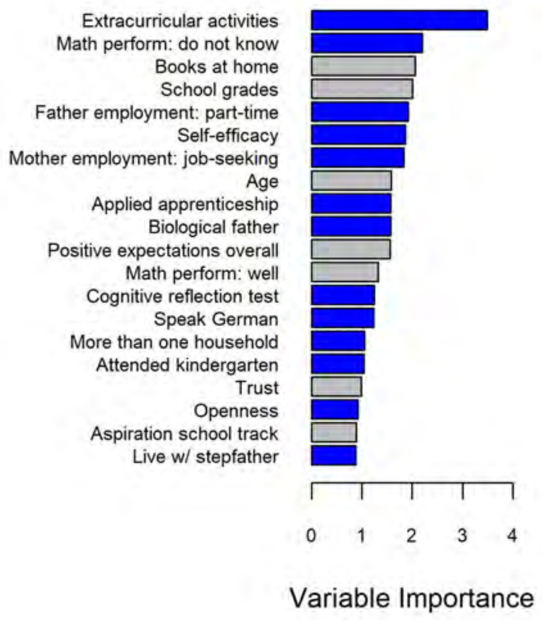
Figures and Tables

Figure 3.1: ROC Curves

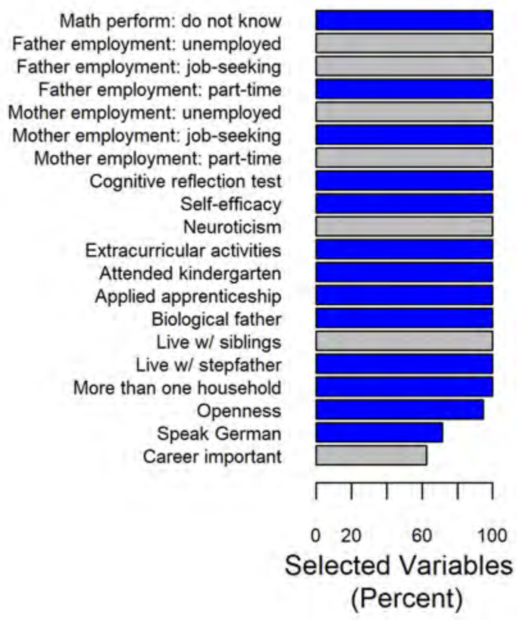


Notes: This figure shows the ROC curves for the LASSO, Post-LASSO, Random Forest, and the logit model.

Figure 3.2: Important Predictors



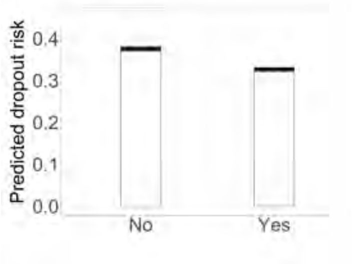
Panel A: Variable Importance from Random Forest



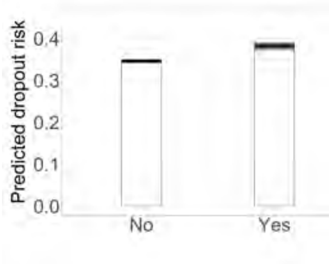
Panel B: Selected Variables from LASSO/Post-LASSO (in Percent)

Notes: Panel A (Random Forest): This figure shows the mean of variable importance of the subsets, each observation used once as a test sample. Panel B (LASSO): This figure shows the frequency of selection of a variable, i.e., the share of 254 subsets in which a variable is selected. 100 indicates that a variable is selected in all subsets, i.e., in 100 percent of the subsets. Only variables with non-zero selection probability are listed. Selected city identifiers are not shown. Variables shaded in blue are those that are among the top 20 variables in both models.

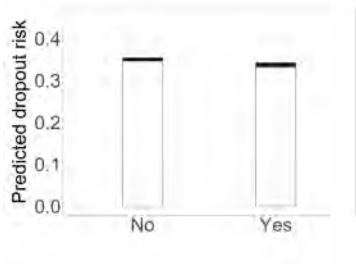
Figure 3.3: Partial Effects of Six Selected Variables (Random Forest)



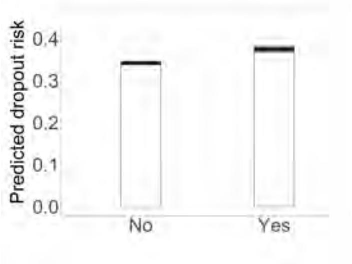
Panel A: Extracurricular Activities



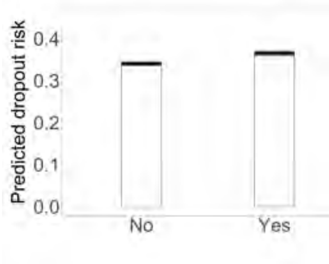
Panel B: Mother Employment (Job-Seeking)



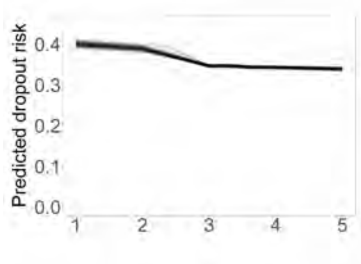
Panel C: Father Employment (Part-Time)



Panel D: Math Performance Do Not Know



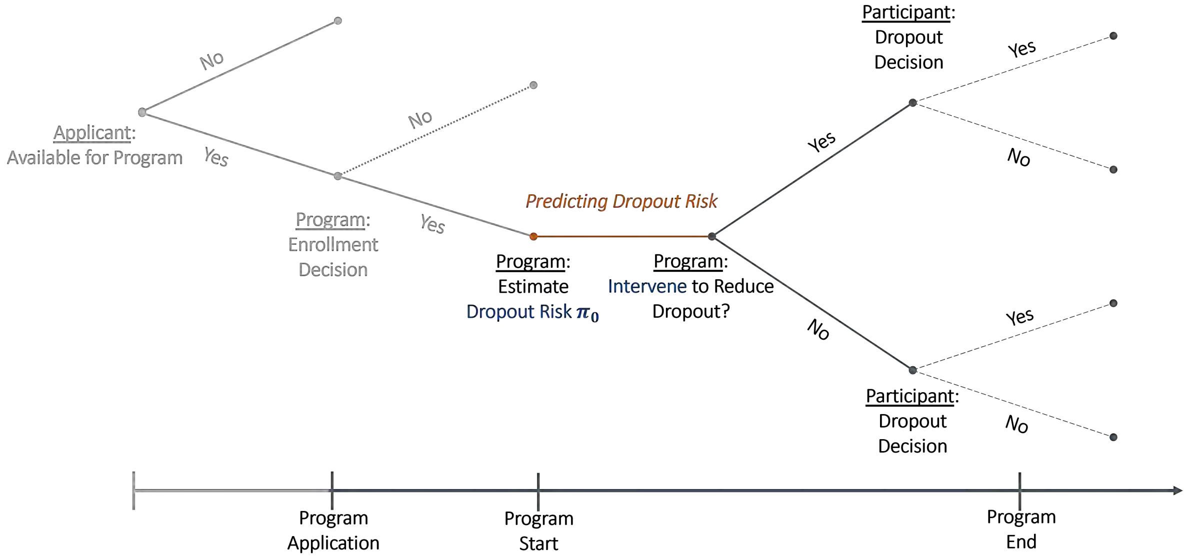
Panel E: Applied for Apprenticeship



Panel F: Self-Efficacy

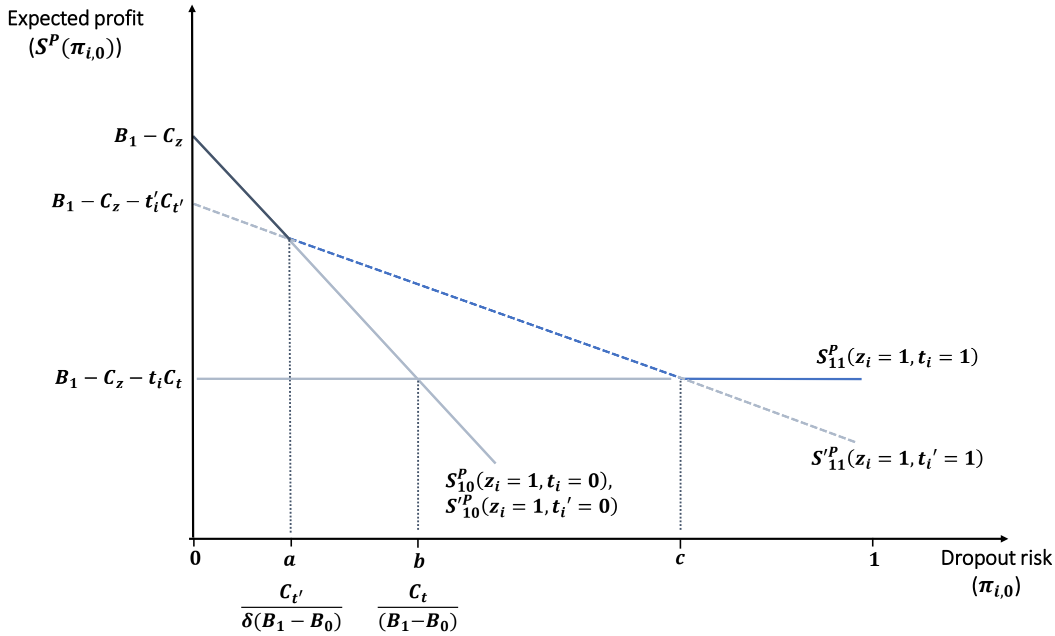
Notes: This figure shows the partial dependence plots from Random Forest for the six most important variables with the highest importance score in the Random Forest model and selected in each round by LASSO.

Figure 3.4: The Role of Predicting Dropout and the Decisions Related to Program Dropout



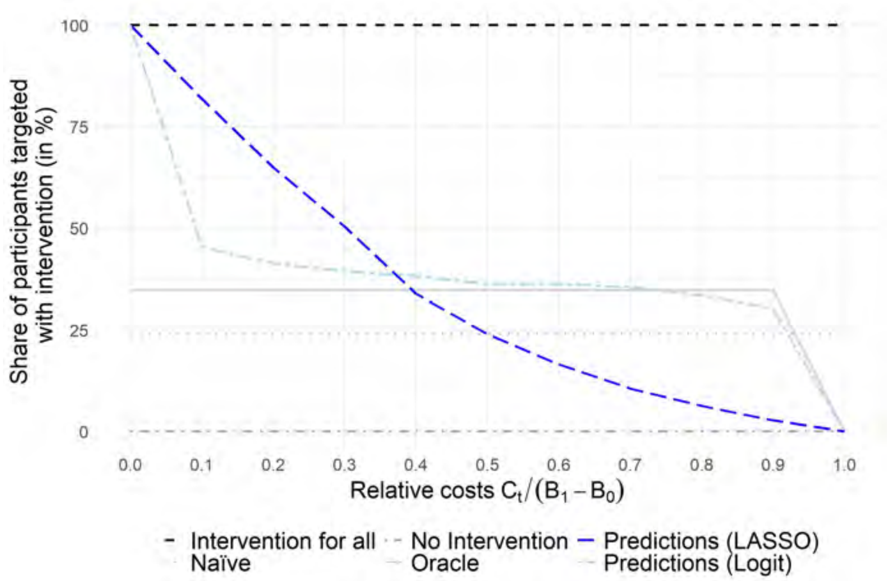
Notes: This figure outlines the choices of a program participant and the program agency related to program dropout as analyzed in this paper. After enrollment, the program agency forms expectations about the individual dropout risk of each participant. The initial selection (in light grey) inherent for every social program is assumed as exogenously given.

Figure 3.5: Illustration of Intervention Choice for t_i and t_i' as a Function of Dropout Risk



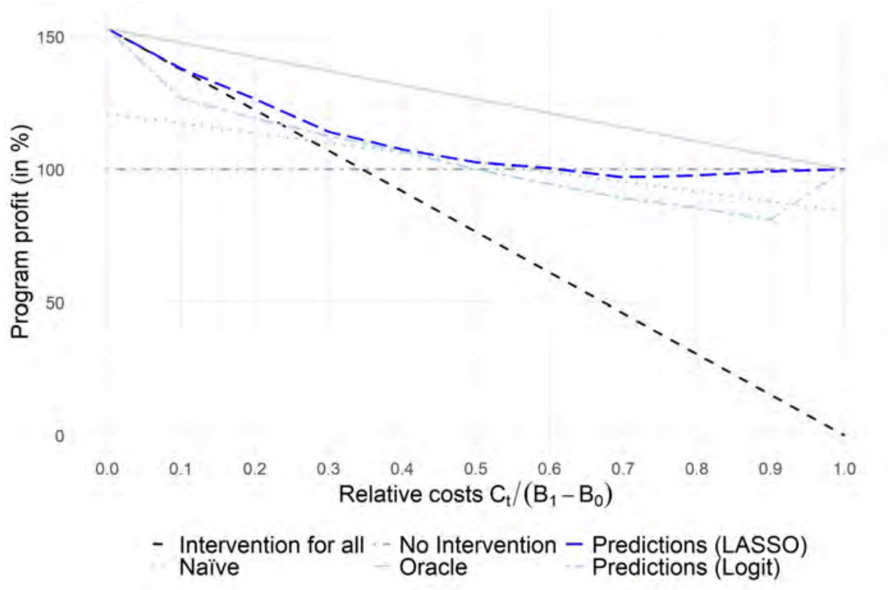
Notes: This figure illustrates the relationship between a participant's individual dropout risk and the program agency's expected profit for a binary intervention t_i and/or t_i'. The figure illustrates the case in which the less effective intervention t_i' has lower costs than $t_i (C_{t'} < C_t)$.

Figure 3.6: Share of Participants Targeted with Additional Intervention



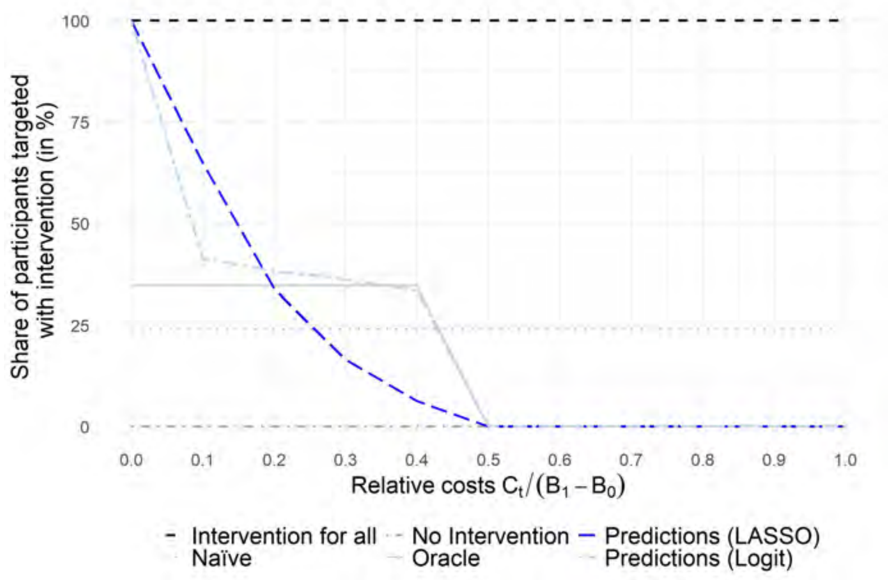
Notes: This figure shows the share of participants targeted with interventions depending on the associated relative costs. Normalized values for $B_1 = 1$ and $B_0 = 0$.

Figure 3.7: Expected Program Profit



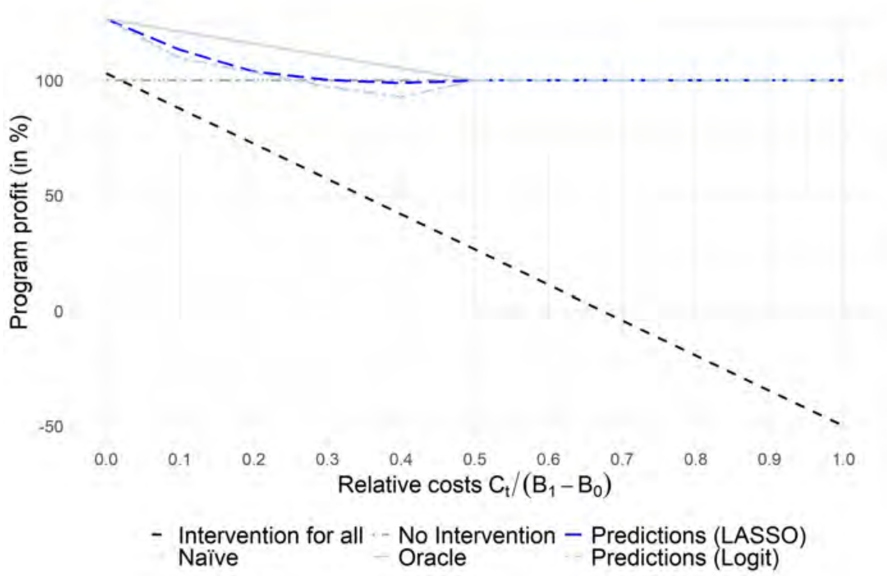
Notes: This figure shows the expected program profit depending on the associated relative costs. Normalized values for $B_1 = 1$ and $B_0 = 0$. Expected program profit is calculated according to Table 3.7.

Figure 3.8: Share of Participants Targeted with Additional Intervention (Partial Reduction of Dropout Risk)



Notes: This figure shows the share of participants targeted with an intervention depending on the associated relative costs under the assumption that interventions reduce dropout risk by $\delta = 0.5$. Normalized values for $B_1 = 1$ and $B_0 = 0$.

Figure 3.9: Expected Program Profit (Partial Reduction of Dropout Risk)



Notes: This figure shows the expected program profit depending on the associated relative costs under the assumption that interventions reduce dropout risk by δ = 0.5. Normalized values for $B_1 = 1$ and $B_0 = 0$. Expected program profit is calculated according to Appendix Table A3.3.

Table 3.1: Descriptive Statistics of the Disadvantaged Adolescents in the Study Sample

	Mean (1)	Std. Dev. (2)	Germany (3)	p-value (1-3) (4)	Germany w/o highest track (5)	p-value (1-5) (6)
Baseline characteristics						
Male	0.43	0.50	0.49	0.03	0.52	0.00
Migrant	0.58	0.50	0.36	0.00	0.38	0.00
Mother employment (part- or full-time)	0.59	0.49	0.80	0.00	0.78	0.00
Father employment (part- or full-time)	0.75	0.44	0.94	0.00	0.93	0.00
Books (less than 25)	0.45	0.50	0.18	0.00	0.25	0.00
Fail math	0.15	0.35	0.06	0.00	0.07	0.00
Fail German	0.11	0.31	0.01	0.00	0.02	0.00
Extraversion (Big5)	-0.08	0.96	0.00	0.18	-0.02	0.34
Agreeableness (Big5)	-0.03	1.21	0.00	0.74	0.02	0.56
Conscientiousness (Big5)	0.09	0.98	0.00	0.15	0.06	0.67
Openness (Big5)	-0.01	0.97	0.00	0.82	-0.04	0.64
Neuroticism (Big5)	0.17	0.94	0.00	0.01	0.02	0.01
Extracurricular activities	0.46	0.50	0.76	0.00	0.71	0.00

Notes: The RYL sample consists of 254 participants of the evaluation study who were assigned to participate in the mentoring program. The NEPS sample consists of 11,768 observations of the start cohort 4, starting in grade 9 with non-missing information on the relevant variables. The sample "Germany w/o highest track" in column (5) comprises only students from the lower track schools in Germany, i.e., without the *Gymnasium*. Mentees report their grades in math and German in the survey according to German standard notation (i.e., 1 = best, 6 = worst) and those receiving grades 5 or 6 are coded as failing the subject. Some students in each of the sample did not receive grades in the respective subject. Values for the Big Five personality traits are standardized with the German NEPS sample. The t-statistic is calculated with the following formula: $t = (\text{mean}_{\text{NEPS}} - \text{mean}_{\text{RYL}})/\sqrt{((\text{std}^2_{\text{NEPS}})/(\text{sample size}_{\text{NEPS}}) + (\text{std}^2_{\text{RYL}})/(\text{sample size}_{\text{RYL}}))}$ and p-values are then calculated with this t-statistic and the respective degrees of freedom (two-sided).

Table 3.2: Reasons for Dropout

Reasons	Mentee	Mentor
Mentee/mentor did not have enough time	48.86	25.81
Mentee/mentor no longer felt like it	40.91	0.00
Mentee no longer needed help	27.27	22.58
Mentor/mentee has not been in touch anymore	16.85	48.39
Mentee/mentor expected something different from mentoring	13.64	9.68
Mentee/mentor did not get along well with mentor/mentee	9.09	9.68
Mentor could not help	7.95	25.81
Mentor moved to another city	6.82	6.45
We achieved everything that we planned to do	3.41	0.00
We both did not get along well	3.41	3.23
Mentee's parents have forbidden me to participate	2.27	3.23
Mentor no longer attends university	0.00	0.00
Mentee has left school	–	6.45
Mentee moved to another city	–	3.23

Notes: The mentee sample consists of 88 participants of the evaluation study who were assigned a mentor and who indicated that their relationship had ended. The mentor sample consists of 114 university students among which 31 indicate that the relationship does not exist anymore.

Table 3.3: Pre-Selection of Predictors from the Baseline Survey

Categories	Variables
Socio-economic predictors	Gender (male), age, speak German, migration background, first-generation migrant, second-generation migrant
Parental background	More than one hh (household), live with mother, live with father, live with stepfather, live with stepmother, live with siblings, live with grandparents, mother employment, father employment, mother university education, father university education, biological father, biological mother, mother speaks German, father speaks German, parental support homework, books at home
Life circumstances	Aspiration school degree, knows occupational career, career important, school and future important
Help from others	Private teaching
Lack of orientation in career choice	No plan after school, not sure about occupational career
Applications	Already applied for apprenticeships
School & kindergarten	Kindergarten attendance, school hours missed, relative math performance, self-reported school grades, school hours, satisfied with school, classmates ambitious, classmates do not care, classmates expect effort
Social life	Extracurricular activities, meet friends, number of friends
Preferences	Trust, risk
Non-cognitive skills	Big5 (extraversion, neuroticism, openness to experience, conscientiousness, agreeableness), internal locus of control, external locus of control, self-efficacy
Personality traits	Confidence, satisfaction, patience, prosociality
Tests	Cognitive reflection test, effort test
Future outlook	Life well, become unemployed, career successful, make money, get apprenticeship

Notes: Pre-selected variables from the baseline survey.

Table 3.4: Realizations and Predictions of Dropout

		Prediction	
		Dropout ($\widehat{W}_i = 0$)	No Dropout ($\widehat{W}_i = 1$)
Realization	Dropout ($W_i = 0$)	True Positives $W_i = 0, \widehat{W}_i = 0$	False Negatives $W_i = 0, \widehat{W}_i = 1$
	No Dropout ($W_i = 1$)	False Positives $W_i = 1, \widehat{W}_i = 0$	True Negatives $W_i = 1, \widehat{W}_i = 1$

Notes: Matrix representing the actual and predicted situation of dropout and no dropout.

Table 3.5: Performance Measures of Algorithms

	Random Forest	Post-LASSO	LASSO	Logit
Accuracy (>0.5)	0.6535	0.6929	0.6890	0.6535
Sensitivity (Recall) (>0.5)	0.0795	0.5000	0.3977	0.5227
Precision (>0.5)	0.5000	0.5641	0.5641	0.5000
AUC	0.6405	0.7426	0.7221	0.6753

Notes: Performance measures of the four algorithms Random Forest, Post-LASSO, adaptive LASSO and logit. Threshold = 0.5 for probability to be class = 1. Test-Sample (Leave-One-Out). Out of 89 variables in total, seven variables dropped in logit: migrant first generation, city identifier MA, mother employment (do not know), father employment (do not know), mother university education (do not know), father university education (do not know), math performance (do not know). Number of variables chosen by LASSO is between 24 and 31.

Table 3.6: Differences in Baseline Characteristics between Dropouts and Non-Dropouts

	Non-Dropouts $W_i = 1$		Dropouts $W_i = 0$		$E[X\|W_i = 1] - E[X\|W_i = 0]$	
Baseline variable (X)	Mean	Std. Dev.	Mean	Std. Dev.	Difference	p-value
Books at home	3.02	1.58	2.41	1.27	0.61	<0.00
Mother employment (job-seeking)	0.05	0.23	0.15	0.36	-0.10	0.03
Self-reported grades	4.03	0.90	3.78	0.97	0.25	0.05
Agreeableness (Big5)	3.42	0.82	3.46	0.76	-0.04	0.74
Neuroticism (Big5)	2.95	0.78	2.84	0.85	0.11	0.35
Openness (Big5)	3.55	0.87	3.31	1.02	0.24	0.06
Conscientiousness (Big5)	3.25	0.83	3.15	0.90	0.10	0.36
Extraversion (Big5)	3.35	0.86	3.41	0.85	0.06	0.56
Cognitive reflection test	0.61	0.79	0.41	0.59	0.20	0.02
Effort test	3.15	0.98	3.13	0.86	0.02	0.88
Prosociality	8.51	1.33	8.31	1.68	0.06	0.32
Male	0.45	0.50	0.38	0.49	0.07	0.24
Migrant	0.54	0.50	0.66	0.48	-0.11	0.06
Math performance well	0.43	0.50	0.27	0.45	0.16	0.01
Extracurricular activities	0.54	0.50	0.29	0.46	0.25	0.00

Notes: Selected baseline characteristics between participants who stayed in the program and those who dropped out early. p-values reported for the difference in means. The usual German ordering of grades is reversed such that self-reported grades are shown in reversed order, i.e., higher values indicate better achievement.

Table 3.7: Expected Program Profit as a Function of B_0, B_1, and C_t

		Program Participation	
		W_i = 0 (Dropout)	W_i = 1
Intervention	t_i^* = 0 if $\hat{\pi}_{i,0} \leq C_t/(B_1 - B_0)$	B_0	B_1
Choice	t_i^* = 1 if $\hat{\pi}_{i,0} > C_t/(B_1 - B_0)$	$B_1 - C_t$	$B_1 - C_t$

Notes: Expected program profit derived from the maximization problem (3.4) described in section 3.5.2.

Table 3.8: Overview of the Five Scenarios

Scenario	Predictions	Decision Rule Intervene if
Predictions (ML & logit)	$\hat{\pi}_{i,0}$ from ML algorithm & logit	$\hat{\pi}_{i,0} > C_t/(B_1 - B_0)$
Oracle	$\hat{\pi}_{i,0} = \pi_{i,0}$	$\hat{\pi}_{i,0} > C_t/(B_1 - B_0)$
Naïve	$\hat{\pi}_{i,0}$ from ML algorithm	$\hat{\pi}_{i,0} > 0.5$
Intervention for all	$\hat{\pi}_{i,0} = 1 \forall i$	$\hat{\pi}_{i,0} \geq 0$
No intervention	–	never

Notes: Description of the five scenarios used to calculate the number of participants and expected program profit.

Table 3.9: Performance Measures for Different Cut-Offs

Cut-off	Precision		Sensitivity	
	LASSO	Logit	LASSO	Logit
0.1	0.404	0.487	0.955	0.636
0.2	0.467	0.505	0.875	0.602
0.3	0.484	0.510	0.705	0.580
0.4	0.547	0.515	0.534	0.568
0.5	0.574	0.500	0.398	0.523
0.6	0.619	0.500	0.295	0.523
0.7	0.519	0.500	0.159	0.511
0.8	0.563	0.518	0.102	0.500
0.9	0.714	0.494	0.057	0.432

Notes: Table shows the two performance measures precision and sensitivity for different cut-offs to classify participants as dropouts or non-dropouts.

Appendix

Appendix A3.1 Machine Learning Algorithms

This section describes the basics of classification trees. It also briefly introduces other algorithms that we use to predict dropout (ridge regression and neural networks).

Classification Trees. Classification trees are used for predicting a categorical response (James et al., 2015). To grow a (classification) tree, recursive binary splitting is used, i.e., the data is iteratively split into two regions (branches). At each node *m*, a predictor and a cut point are selected such that splitting the predictor space into regions leads to 'good' splits. To evaluate the quality of a particular split, we use the Gini impurity as the loss function which is "a measure of node purity" (James et al. 2015, p. 312). Gini impurity describes the probability of incorrectly classifying a randomly chosen element in the dataset if we labeled it randomly in line with the class distribution in the dataset. The best and lowest possible impurity is zero. We can weigh the impurity of each branch by the number of elements it has, and then calculate the amount of impurity removed (i.e., subtract the weighted impurities from the original impurity), which is the Gini Gain. The best split is obtained by maximizing the Gini Gain (Zhou, 2019). The Gini impurity is mathematically defined by the following equation:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}),$$

where $\hat{p}_{mk}$ is the proportion of training observations in the m^{th} region that are from the k^{th} class.

The same prediction is made for every observation that falls into one region: the prediction is made according to the most commonly occurring class of observations in the training data in that region to which an observation belongs (James et al., 2015). In our case, the classification is either zero or one since we have a binary dependent variable.

Classification decision trees have the advantage of being easy to interpret. However, single trees usually perform worse in terms of predictive accuracy than other classification approaches and are often not robust, i.e., small changes in the data can result in large changes in the estimated tree (James et al., 2015).

Ridge Regression. Ridge regression works similar as LASSO. Compared to LASSO, ridge shrinks all coefficients towards zero but does not set any of them to zero (James et al., 2015). While LASSO uses an $L1$-penalty, i.e., the sum of the coefficients' absolute values, ridge uses an $L2$-penalty which is equal to the sum of the square of the coefficients ($\lambda \sum_{j=1}^{p} \beta_j^2$).

Neural Networks. Neural networks are a type of ML that is inspired by the human nervous system (Aggarwal, 2018). The main concept involves extracting linear combinations of input variables to create derived features, which are then used to model the target variable as a nonlinear function (Hastie et al., 2016). A neural network is comprised of interconnected layers of artificial neurons, referred to as nodes or units, which receive input signals, perform computations, and generate output signals. During the training phase, neural networks learn by adapting the weights assigned to the connections between neurons. As learning mechanism, we use backpropagation which optimizes the weights through minimization of the loss function. The loss function involves comparing the network's predictions with the actual target values. We use a single layer fully connected neural network where a set of inputs is directly linked to an output without the presence of additional hidden layers where every connection has an individual weight.

As the activation function, which introduces nonlinearity to the network, we use the sigmoid function. Since we predict a probability of a binary class, the sigmoid function is the preferred choice (Aggarwal, 2018).

Appendix A3.2 Model Appendix

Enrollment Decision. The decision to offer the program depends on the comparison of the expected benefits $E[B^P]$ and the costs C_z. The maximization problem of the program agency reads:

$$\max_{z_i \in \{0,1\}, \forall i} B_0 + z_i(1 - \pi_{i,0})(B_1 - B_0) - z_i C_z. \tag{B.1}$$

The agency decides to offer the program ($Z_i = 1$) if the expected benefits exceed the costs, i.e., $(1 - \pi_{i,0})(B_1 - B_0) > C_z$, and declines program access otherwise ($Z_i = 0$).

The Two-Stage Model. The selection of applicants and interventions through additional measures can also be modeled in a two-stage model. With additional measures available to reduce dropout, this directly influences the decision to enroll an applicant in the first place. The optimal intervention policy t_i^* determines the continuation value if program access is granted ($Z_i = 1$) or not. The maximization problem that determines both the optimal enrollment choice $Z_i = z_i^*$ and the optimal intervention choice t_i^* reads as follows:

$$\max_{z_i \in \{0,1\}, t_i \in \{0,1\}, \forall i} B_0 + z_i(1 - \pi_{i,0} + t_i \pi_{i,0})(B_1 - B_0) - z_i C_z - t_i C_t. \tag{B.2}$$

The profit function is then defined as the program profit given the optimal program choices, $S^{P*} \equiv S^P(t_i^*, z_i^*)$.[1] Comparing the surplus of all available policies

$$(z_i, t_i) \in \{(0, 0); (1, 0); (1, 1)\}$$

results in the following conditions to describe the optimal behavior of the program agency.

The agency will <u>deny</u> program access ($z_i^* = 0$ and $t_i^* = 0$) if:

- *either* the enrollment costs are prohibitively high ($C_z > B_1 - B_0$),
- *or* the enrollment costs are lower than the potential benefit of the program ($C_z < B_1 - B_0$) but still higher than the potential benefit net of the intervention costs ($C_z > B_1 - B_0 - C_t$) such that a later intervention is not appealing and the chances of remaining in the program without an intervention are sufficiently low ($1 - \pi_{i,0} < C_t/(B_1 - B_0)$) such that enrollment is not appealing.

The agency will <u>allow</u> program access but will <u>not intervene</u> ($z_i^* = 1$ and $t_i^* = 0$) if:

- chances of remaining in the program without an intervention are sufficiently high ($1 - \pi_{i,0} > C_z/(B_1 - B_0)$) such that enrolling is still appealing *and* the costs of intervening are higher than the expected increase in the benefits ($C_t > \pi_{i,0}(B_1 - B_0)$).

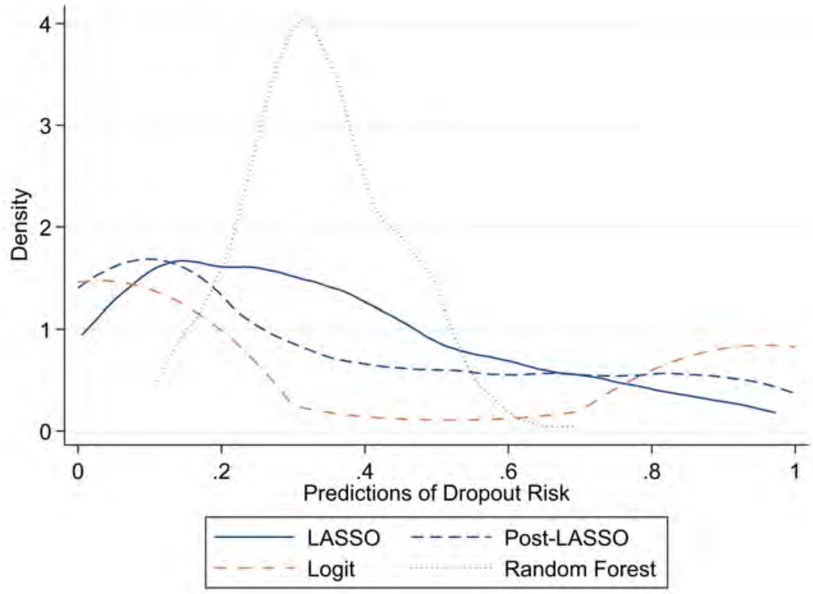
[1] As enrollment is essential, the program agency can use backward induction to determine first the optimal intervention policy t_i^* (that would only be implemented if $Z_i = 1$), and afterwards the optimal enrollment choice $z_i^* | t_i^*$.

The agency will allow program access and intervene ($z_i^* = 1$ and $t_i^* = 1$) if:

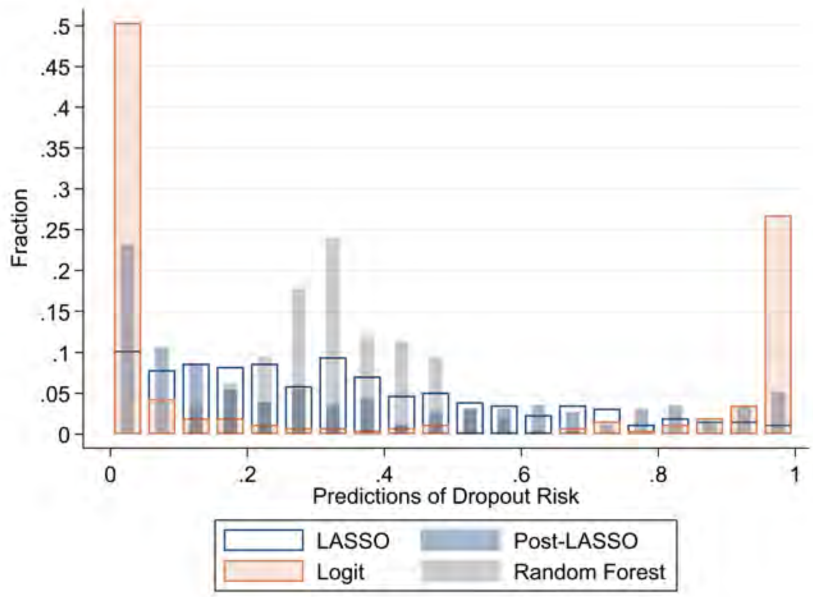
- costs of intervening are lower than the expected increase in the benefits $(C_t < \pi_{i,0}(B_1 - B_0))$ and lower than the potential program benefit net of the enrollment costs $(C_t < B_1 - B_0 - C_Z)$.

Appendix A3.3 Appendix Figures and Tables

Figure A3.1: Distribution and Histogram of Predictions



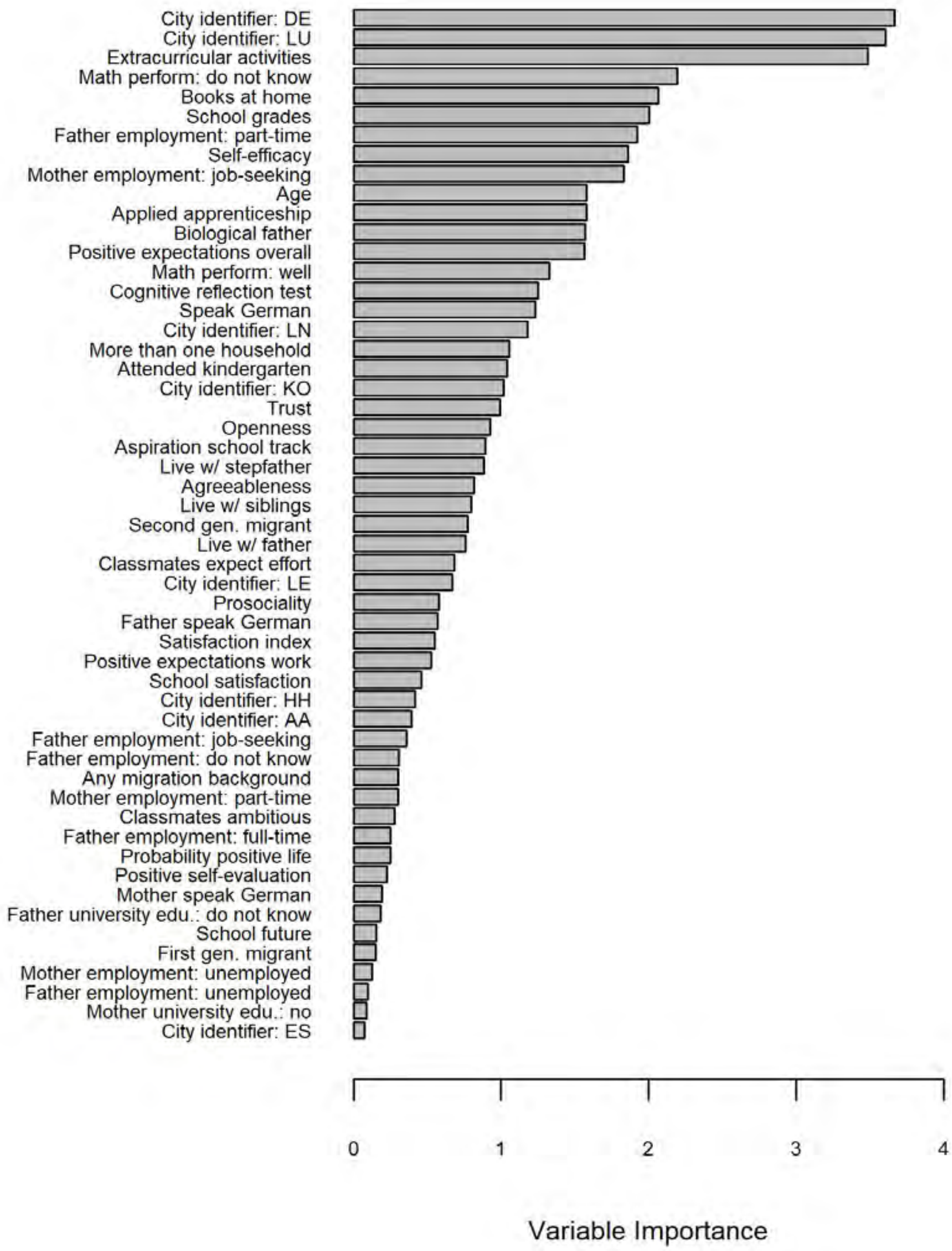
Panel A: Density of Predictions of Algorithms



Panel B: Histogram of Predictions of Algorithms

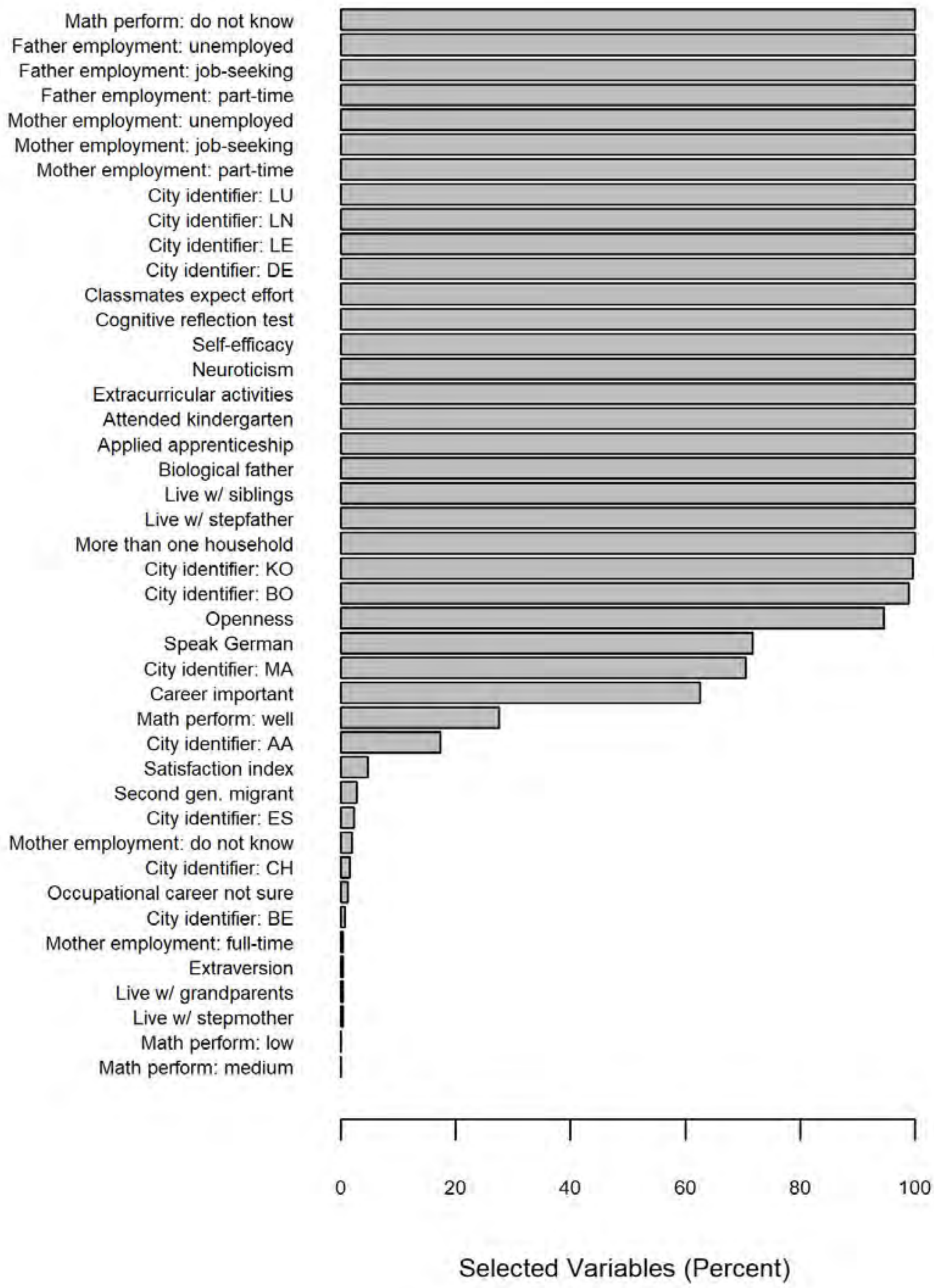
Notes: Panel A shows the density of the predictions from the four algorithms. Panel B shows a histogram of the predictions of the four algorithms with a bin width of five percent.

Figure A3.2: Variable Importance Random Forest (All Predictors)



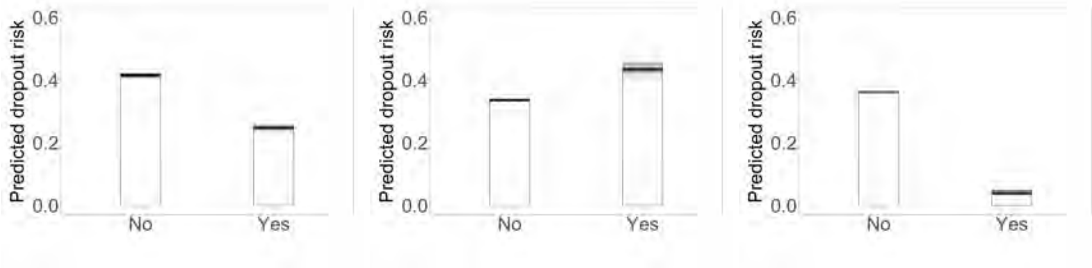
Notes: This figure shows the variable importance according to the Random Forest for all variables with an importance score larger than zero (including city identifiers).

Figure A3.3: Selected Variables from LASSO (All Predictors) in Percent



Notes: This figure shows the frequency of selection of a variable, i.e., the share of 254 subsets in which a variable is selected. 100 indicates that a variable is selected in all subsets, i.e., in 100 percent of the subsets. The figure shows all selected variables from the LASSO model that are selected at least once (including city identifiers).

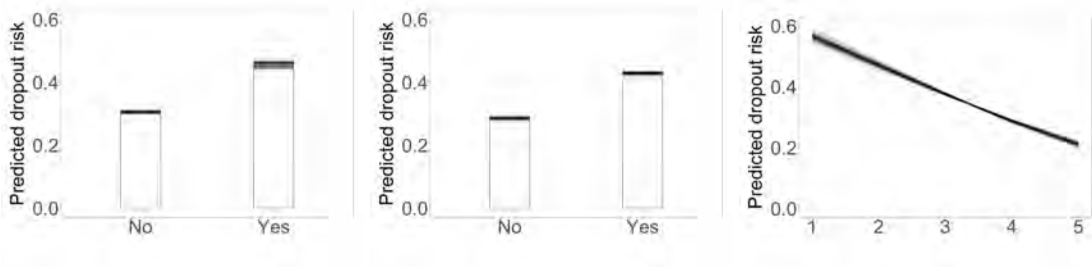
Figure A3.4: Partial Effects of Six Selected Variables (Post-LASSO)



Panel A: Extracurricular Activities

Panel B: Mother Employment (Job-Seeking)

Panel C: Father Employment (Part-Time)



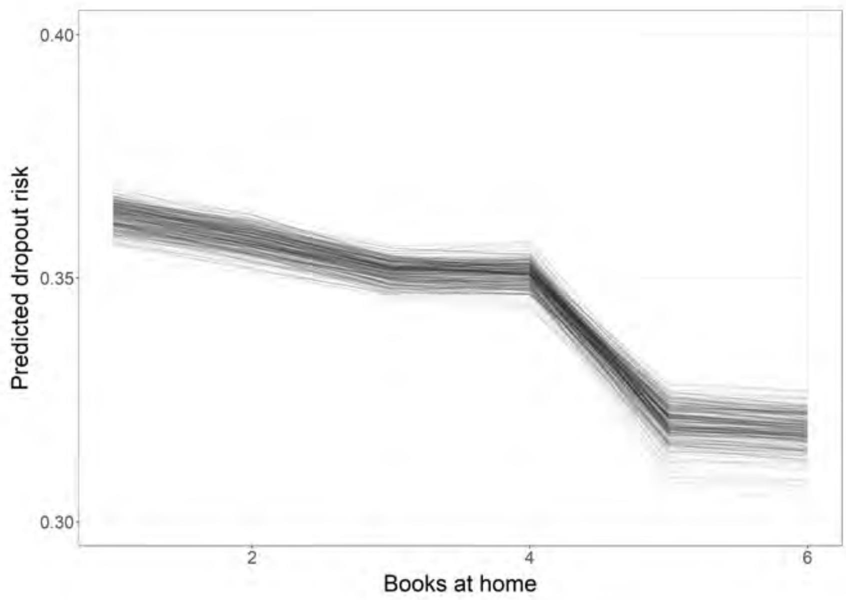
Panel D: Math Performance Do Not Know

Panel E: Applied for Apprenticeship

Panel F: Self-Efficacy

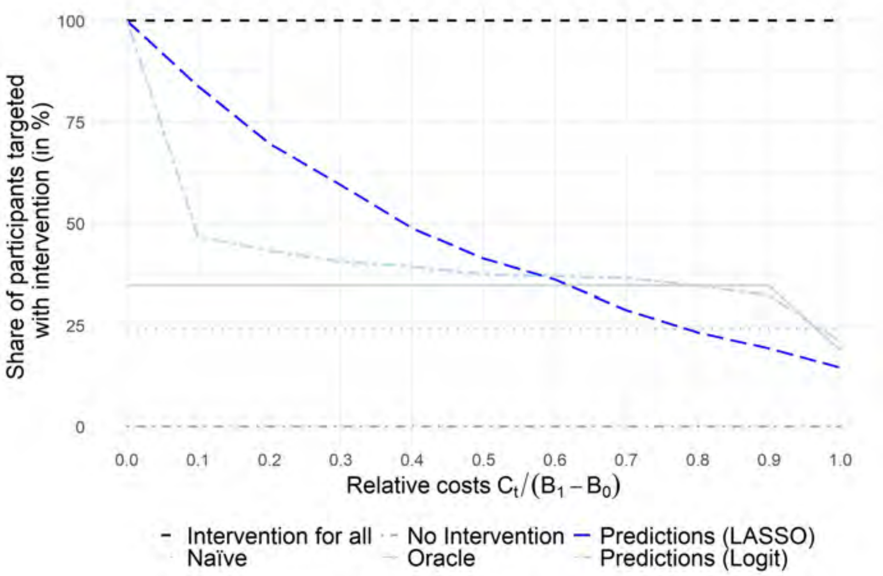
Notes: This figure shows the partial dependence plots from Post-LASSO for the six most important variables with the highest importance score in the Random Forest model and selected in each subset by LASSO.

Figure A3.5: Partial Effects of "Books at Home"



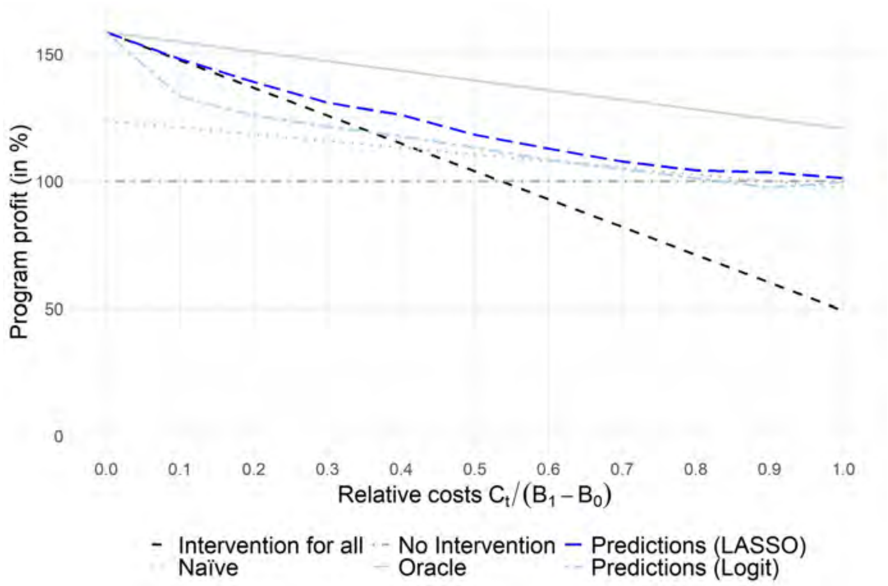
Notes: This figure shows the partial dependence plots from Random Forest for the variable "books at home".

Figure A3.6: Share of Participants Targeted with Additional Intervention (Differential Benefits)



Notes: This figure shows the share of participants targeted with interventions depending on the associated relative costs under the assumption that low-SES mentees yield greater benefits. Normalized values for $B_1 = 1$ for higher-SES mentees, $B_1 = 2$ for low-SES mentees, and $B_0 = 0$.

Figure A3.7: Expected Program Profit (Differential Benefits)



Notes: This figure shows the expected program profit depending on the associated relative costs under the assumption that low-SES mentees yield greater benefits. Normalized values for $B_1 = 1$ for higher-SES mentees, $B_1 = 2$ for low-SES mentees, and $B_0 = 0$. Expected program profit calculated according to Table 3.7.

Table A3.1: Missing Values

Variable	No. of missings	Share of missings
Male	0	0.00
Migrant	0	0.00
First generation migrant	0	0.00
Second generation migrant	0	0.00
Age	0	0.00
Speak German	0	0.00
More than one household	0	0.00
Live with father	0	0.00
Live with mother	0	0.00
Live with stepfather	0	0.00
Live with stepmother	0	0.00
Live with siblings	0	0.00
Live with grandparents	0	0.00
Mother employment	3	0.01
Father employment	7	0.03
Father university education	2	0.01
Mother university education	3	0.01
Biological mother	1	<0.01
Biological father	4	0.02
Parental support homework	0	0.00
Books at home	0	0.00
Father speaks German	5	0.02
Mother speaks German	3	0.01
Aspiration school degree	1	<0.01
Knows occupational career	3	0.01
Career important	0	0.00
School important	0	0.00
Private teaching	0	0.00
No plan after school	22	0.09
Not sure about occupational career	22	0.09
Applied for apprenticeship	5	0.02
Kindergarten attendance	0	0.00
School hours missed	5	0.02
Relative math performance	0	0.00
Self-reported grades	46	0.18
Satisfied school	0	0.00
School hours	4	0.02
Classmates ambitious	0	0.00
Classmates do not care	2	0.01
Classmates expect effort	2	0.01

(continued on next page)

Table A3.1 (continued)

Variable	No. of missings	Share of missings
Extracurricular activities	0	0.00
Meet friends	1	<0.01
No. of friends	2	0.01
Trust	4	0.02
Risk	2	0.01
Extraversion (Big5)	0	0.00
Neuroticism (Big5)	0	0.00
Openness to experience (Big5)	0	0.00
Conscientiousness (Big5)	0	0.00
Agreeableness (Big5)	1	<0.01
Internal locus of control	1	<0.01
External locus of control	3	0.01
Self-efficacy	2	0.01
Confidence	1	<0.01
Satisfaction	0	0.00
Patience	1	<0.01
Prosociality	0	0.00
Cognitive reflection test	24	0.09
Effort test	31	0.12
Life well	0	0.00
Get apprenticeship	0	0.00
Become unemployed	0	0.00
Career success	3	0.01
Make money	0	0.00

Notes: Table shows the number and share of missing values for each variable used in the analyses.

Table A3.2: Performance Measures of Further Algorithms

	Normal LASSO	Ridge regression	Neural networks
Accuracy (>0.5)	0.6693	0.6535	0.6811
Sensitivity (Recall) (>0.5)	0.3409	0.2727	0.2500
Precision (>0.5)	0.5357	0.5000	0.5946
AUC	0.5921	0.6972	0.5798

Notes: Performance measures of three additional algorithms (Normal LASSO, ridge regression, and neural networks). Threshold = 0.5 for probability to be class = 1. Test-Sample (Leave-One-Out). Ridge regression and neural networks are described in Appendix A3.1.

Table A3.3: Expected Program Profit as a Function of B_0, B_1, and C_t (Partial Reduction of Dropout Risk)

		Program participation	
		W_i = 0 (Dropout)	W_i = 1
Intervention	$t_i'^*$ = 0 if $\delta\hat{\pi}_{i,0} \leq C_{t'}/(B_1 - B_0)$	B_0	B_1
Choice	$t_i'^*$ = 1 if $\delta\hat{\pi}_{i,0} > C_{t'}/(B_1 - B_0)$	$\delta B_1 + (1-\delta)B_0 - C_{t'}$	$B_1 - C_{t'}$

Notes: Expected program profit derived from maximization problem in equation (3.6) described in section 3.5.2.

Table A3.4: Number of Participants Targeted with Additional Intervention

	Costs of intervention relative to benefits $C_t/(B_1 - B_0)$										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Predictions Scenario											
Intervene if $\hat{\pi}_{i,0}$	> 0	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	> 0.6	> 0.7	> 0.8	> 0.9	never
$\hat{\pi}_{i,0}$ *calculated with ...*											
Random Forest	254	254	228	152	62	14	1	0	0	0	0
Post-LASSO	254	168	132	108	88	78	62	46	35	22	0
LASSO	254	208	165	128	86	61	42	27	16	7	0
Logit	254	115	105	100	97	92	92	90	85	77	0
Oracle scenario											
$\hat{\pi}_{i,0} = \pi_{0,i}$	88	88	88	88	88	88	88	88	88	88	0
No intervention scenario											
No predictions	0	0	0	0	0	0	0	0	0	0	0
Intervention for all scenario											
$\hat{\pi}_{i,0} = 1 \forall i$	254	254	254	254	254	254	254	254	254	254	254
Naïve scenario											
Intervene if $\hat{\pi}_{i,0}$	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5
LASSO	61	61	61	61	61	61	61	61	61	61	61

Notes: Number of adolescents out of overall 254 participants that should optimally be targeted with an intervention according to equation (3.5). A short explanation of the scenarios can be found in Table 3.8.

Table A3.5: Agency's Expected Profit

	Costs of intervention relative to benefits $C_t/(B_1 - B_0)$										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Predictions Scenario											
Intervene if $\hat{\pi}_{i,0}$	> 0	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	> 0.6	> 0.7	> 0.8	> 0.9	never
$\hat{\pi}_{i,0}$ *calculated with ...*											
Random Forest	254.0	228.6	206.4	187.4	167.2	166.0	165.4	166.0	166.0	166.0	166.0
Post-LASSO	254.0	227.2	208.6	191.6	177.8	171.0	163.8	160.8	158.0	158.2	166.0
LASSO	254.0	229.2	210.0	189.6	178.6	170.5	166.8	161.1	162.2	164.7	166.0
Logit	254.0	210.5	198.0	187.0	177.2	166.0	156.8	148.0	142.0	134.7	166.0
Oracle scenario											
$\hat{\pi}_{i,0} = \pi_{0,i}$	254.0	245.2	236.4	227.6	218.8	210.0	201.2	192.4	183.6	174.8	166.0
No intervention scenario											
No predictions	166.0	166.0	166.0	166.0	166.0	166.0	166.0	166.0	166.0	166.0	166.0
Intervention for all scenario											
$\hat{\pi}_{i,0} = 1 \forall i$	254.0	228.6	203.2	177.8	152.4	127.0	101.6	76.2	50.8	25.4	0.0
Naïve scenario											
Intervene if $\hat{\pi}_{i,0}$	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5
LASSO	201.0	195.0	189.0	183.0	177.0	171.0	164.0	158.0	152.0	146.0	0.0

Notes: Expected program profit for 254 observations. Normalized values for $B_1 = 1$ and $B_0 = 0$. Expected profit calculated according to Table 3.7. A short explanation of the scenarios can be found in Table 3.8.

4 Luck or Effort: Perceptions of the Role of Circumstances in Education and the Demand for Targeted Spending*

4.1 Introduction

It is well documented that students' success correlates strongly with parental background, leading to educational inequality by socio-economic status (SES) of parents (e.g., Schütz et al., 2008; Björklund and Salvanes, 2011; OECD, 2018). As better educational attainment is also rewarded with higher wages on the labor market (e.g., Card, 1999), unequal chances of students from different parental backgrounds can have severe implications for economic inequality and inequality of opportunity (e.g., Nickell, 2004; Corak, 2013). Possible ways to improve educational outcomes of students from less advantaged backgrounds include redistributive measures and targeted support, either from (i) public sources, for example redistributive education spending or (ii) private sources, for example donations. The feasibility of implementing targeted support crucially depends on public endorsement for redistributive measures. While a large strand of research has explored preferences for governmental redistribution (e.g., Alesina et al., 2018; Hoy and Mager, 2021), we contribute to the literature by providing evidence on spending preferences of survey respondents for both stated preferences for public redistributive education spending and revealed preferences for private donations. Measuring revealed preferences offers a large advantage over conventional survey measures by addressing the common concern that survey measures of preferences for redistribution fail to capture actual behavior and are prone to experimenter demand effects (Haaland et al., 2023).

In this paper, we hypothesize that public preferences for targeted spending depend on the perception of the interplay between parental disadvantage and the education system. Therefore, we investigate how preferences for private donations and for redistributive education spending change when respondents receive information on the differences in educational outcomes of students from more and less advantaged parental backgrounds. In addition, we also examine how these perceptions affect respondents' beliefs regarding the extent to which external circumstances and effort are decisive for educational success: some may attribute success to effort and poor outcomes to failure to seize opportunities, while others may see economic disadvantage as bad luck or external circumstances. We use data from a large-scale survey experiment ($N > 2{,}000$) in a sample of the population aged 18 and older in Germany. At the start of the experiment, we randomly select treatment respondents who receive accurate information on the educational outcomes of 15-year-old students in Germany. For the

* This chapter is co-authored with Elisabeth Grewenig and Katharina Werner.

treatment information, we focus on the difference in the share of students from more and less advantaged families who attend academic-track secondary schools (*Gymnasium*): 49 percent of students from more advantaged families and 19 percent of students from less advantaged families attend an academic-track secondary school, resulting in a 30-percentage point SES gap in academic-track attendance.[1]

The information experiment allows us to investigate the effect of providing accurate information on academic-track attendance rates of students from more and less advantaged backgrounds on respondents' donation decisions and their preferences for increases in redistributive education spending by the government. Importantly, providing information on both more and less advantaged students allows us to study a more nuanced picture of redistribution: treated respondents might become more concerned about the share of students from less advantaged backgrounds who attend academic-track schools, which in turn, could translate into higher demand for targeted support. Alternatively, respondents might also be misinformed about the share of students from more advantaged backgrounds who attend academic-track schools, which could reduce demand for targeted support.

Similarly, it is ex-ante unclear whether treatment effects on private targeted support, i.e., donations, will correspond to effects on public redistributive education spending. If the treatment information increases the concerns for the educational outcomes of students from less advantaged backgrounds, respondents might increase their preferences for targeted spending from both private and public sources. However, previous literature shows mixed results on whether information can change respondents' policy preferences (e.g., Cruces et al., 2013; Kuziemko et al., 2015; Hoy and Mager, 2021). While inertia of public preferences might have a multitude of causes, a common explanation is that policy reforms are unlikely to be seen as an effective solution if the perceived capacity of government is low. Directly measuring donation behavior in addition to preferences for redistributive public spending allows us to investigate whether information changes preferences for targeted spending in a highly transparent environment. Respondents who are faced with a donation decision are well-informed about their own opportunity costs of the donated amount, while the opportunity costs of funds used for public targeted support are likely to be more opaque: in particular, respondents might be unsure whether increased spending on students from less advantaged backgrounds would be diverted from other education spending benefiting more advantaged children, or other public spending.

We find that respondents largely assume that students' own effort determines their educational success. In the uninformed control group, only 17.3 percent of respondents believe that a high educational degree depends on external circumstances, rather than own effort. Information on the differences in academic-track attendance by parental background strongly increases the perception that external circumstances determine educational success, with a

[1] Attendance of academic-track secondary schools is an important education decision in the German context. Children typically choose the secondary school track after four to six years of primary school. Academic-track secondary schools are the most rigorous and the most common way to obtain a university entrance qualification.

share of 29.4 percent of respondents in the treatment group stating this opinion, an increase by 12.1 percentage points or 69.9 percent. These results even persist into a follow-up survey conducted two weeks after the main survey.

Second, we show that the information treatment also increases private donations to charities supporting students from less advantaged backgrounds with school materials or scholarships by 3.3 tokens, compared to 37.5 tokens at baseline. In the control group, 66.2 percent of respondents donate a positive amount of money. Information on the academic-track school attendance rates increases the share of donators by 9.3 percentage points. Similarly, more respondents decide to donate amounts larger than the control group median. This suggests that the increase in the perceived role of external circumstances rather than students' effort in determining educational success translates into higher demand for private targeted support.

Third, we show that respondents' demand for redistributive education spending by the government remains unchanged by the information treatment. While a large share of respondents (75.1 percent) in the control group supports increased school spending to foster equality of opportunity, the treatment effect of information about academic-track attendance rates of more and less advantaged students is negligible.

To better understand the mechanism of belief updating, we study the degree of misperceptions prior to the information experiment. We also test retention of the treatment information, by comparing respondents' prior beliefs on academic-track attendance rates of students from more and less advantaged backgrounds with re-elicited (posterior) beliefs about these academic-track attendance rates two weeks after the first survey. We describe two key findings: first, we find large misperceptions for all pieces of information at baseline. On average, respondents believe that 71 percent of students from more advantaged families attend academic-track schools (accurate value: 49 percent), while they believe that 30 percent of students from less advantaged families do so (accurate value: 19 percent). This suggests that respondents in the treatment group on average receive a downward information shock for both more and less advantaged students' educational attainment. Second, results from the follow-up survey show that persistent information updating effects are stronger for attendance rates of less advantaged students. Higher retention could imply that treated respondents pay particular attention to information on students from less advantaged backgrounds and might be most concerned about this part of the information treatment. This would be in line with the previously documented treatment effects on the perception that external circumstances determine educational outcomes and increased levels of donations to charities supporting students from less advantaged backgrounds. At the same time, the downward information shock regarding the attendance rates of more advantaged students could temper respondents' demand for public targeted spending if the redistributive effects on more advantaged students are seen as unclear. This highlights a potential role of perceived uncertainty regarding the opportunity costs of additional spending in explaining inertia in public preferences, which we consider an important question for further research.

We explore five alternative reasons why the positive information treatment effects on the perceptions of the importance of circumstances and private donations do not translate into more support for redistributive education spending by the government. We show that respondents think that the concrete policy proposal – increased school spending – is well suited to foster equality of opportunity. Explorative subgroup analyses also reveal that the information treatment does not differentially affect preferences for redistributive education spending of respondents with different educational attainment, political ideologies nor with different levels of trust in the government, suggesting that partisan biases or beliefs about the governmental capability are unlikely to account for the absence of treatment effects. Finally, a more systematic approach using a Causal Forest algorithm similarly fails to detect heterogeneous treatment effects on demand for redistributive education spending along the previously mentioned dimensions as well as a large number of socio-economic characteristics. For private donations, we document suggestive evidence that respondents with lower educational attainment react more strongly to the information treatment.

The remainder of the paper is structured as follows. Section 4.2 introduces the hypotheses that we test in this paper and the related literature. Section 4.3 provides a brief overview of the institutional background of the German education system. Section 4.4 presents the opinion survey, the experimental design, and the estimation strategy. Section 4.5 presents our results. Section 4.6 discusses belief updating, how respondents perceive the information, and potential mechanisms. Section 4.7 concludes.

4.2 Conceptual Framework and Related Literature

There is a long tradition in social sciences to study the relationship between inequality and preferences for redistribution (Piketty, 1995; Bénabou and Ok, 2001; Alesina and Giuliano, 2011; Durante et al., 2014; Almås et al., 2020; Hvidberg et al., 2022).[2] Previous studies show that perceptions over the underlying sources of inequality are an important explanatory factor in determining whether redistribution is favored (Alesina and Glaeser, 2004; Alesina and Angeletos, 2005; Bénabou and Tirole, 2006). On the one hand, individuals might believe that economic success results from effort, and poor outcomes may mainly be attributed to individual failures to use available opportunities. On the other hand, individuals might perceive the system as unfair and assume that economic disadvantage is the result of bad luck or external circumstances beyond someone's control. Because of the different sources attributed to individual success, the two notions might have very different implications for redistribution preferences, with the latter view yielding a higher demand for targeted support for the disadvantaged than the former. Indeed, several papers have empirically confirmed

[2] One strand of this literature has investigated heterogeneities in redistributive preferences using incentivized lab experiments (e.g., Cappelen et al., 2007; Fisman et al., 2007; Cappelen et al., 2013, 2015; Jakiela, 2015; Fisman et al., 2017) while another strand has focused on redistributive preferences in the general population (e.g., Edlund, 1999; Osberg and Smeeding, 2006; Bellemare et al., 2008; Fisman et al., 2015; Falk et al., 2018).

the link between fairness views and distributional preferences using social survey data (e.g., Fong, 2001; Alesina and La Ferrara, 2005; Roth and Wohlfart, 2018). Whether information on differences in economic outcomes between groups gives rise to demands for targeted support might thus depend on the degree to which individuals are seen as responsible for their own economic success, in the sense that success is a consequence of individual effort (e.g., Alesina and Angeletos, 2005; Bénabou and Tirole, 2006). As a result, we would expect demand for redistributive education spending targeted at less advantaged students to increase if the source of unequal outcomes is seen as being less individuals' effort and hence attributed more to external circumstances. In this paper, we test whether information about inequality in educational outcomes in terms of students' academic-track attendance rates by parental SES leads to a change in the perception about the role of circumstances or effort in determining educational success. We thereby extend the literature by providing information about differences in academic-track attendance rates by parental SES while most previous literature has focused on inequality in terms of income. This allows us to understand to which extent respondents interpret the correlation between parental background and academic-track attendance of students as informative on students' individual responsibility for their educational success.

Because people often hold misperceptions about the extent of inequality in society (e.g., Kluegel and Smith, 1986; Norton and Ariely, 2011), we expect the provided information to change average respondent beliefs about academic-track attendance rates of students from different backgrounds in the treatment group compared to the control group. However, it is ex ante unclear which part of the information treatment will be most relevant to respondents: the treatment group receives information on the levels of the share of students attending an academic track for students from advantaged and disadvantaged backgrounds as well as the gap in attendance rates by parental SES. Respondents who are predominantly concerned about the number of students reaching high levels of education regardless of their background, might be most interested in the information on the levels of academic-track attendance. The direction of the effect of respondents' misperceptions about attendance levels is ambiguous. On the one hand, if respondents overestimate the level of students from more advantaged backgrounds who attend the academic-track schools, they may be surprised by the unexpected 'low' share of high-SES students who attend an academic-track school and therefore adjust their perception of the importance of external circumstances for educational success downwards. On the other hand, if respondents previously overestimate the levels of students from less advantaged backgrounds who attend the academic-track schools, respondents may be surprised by the 'low' share of low-SES students who attend an academic-track school and therefore adjust their view about the role of external circumstances for educational success upwards. Respondents might also care about the gap in attendance rates by parental SES if they are mainly worried about social cohesion between groups. The treatment effect might also be ambiguous with respect to the gap in academic-track attendance rates by parental SES, depending on whether respondents care about the absolute difference of the gap or the relative rates. In our paper, we test this hypothesis by examining (i) treatment effects on the perception of the role of circumstances as decisive for educational success, and (ii) respondents' follow-up beliefs about the academic-track attendance levels of high- and low-SES students. Thus, if respondents care more about the share of students from less advantaged backgrounds attending academic-track schools, we expect treated respondents to shift their answer towards reporting a higher importance of external circumstances and to particularly remember this information in a follow-up survey.

Changes in people's perceptions about inequality might in turn affect their willingness to support disadvantaged groups, e.g., through donations to charities. Several papers have analyzed the relationship between inequality and charitable donations. Mostly focusing on income inequality, this strand of the literature shows ambiguous results: some studies, mainly conducted in the lab, find that increases in income inequality are associated with smaller amounts of charitable contributions (Chan et al., 1996; Buckley and Croson, 2006; Coté et al., 2015; Duquette and Hargaden, 2021). However, observational studies document that increases in income inequality can also lead to larger donations (e.g., Payne and Smith, 2015). In contrast to these studies, we do not investigate the effects of changes in inequality per se but observe how information on the inequality in educational outcomes of students from different backgrounds affects private donations. In addition, we do not elicit preferences for donations in general but focus on charities that aim to create equality of opportunities for students from different backgrounds. Measuring respondents' preferences for private redistribution by their donation decision bears several advantages: it constitutes a one-time decision for respondents, it is easily understandable and executable, and most importantly, the donation decision carries a clear implication, where respondents' contributions are directed towards charities dedicated to help less advantaged students. This incentivized measure is a good way to measure respondents' willingness to pay for support for these children as a 'real stakes' question (Stantcheva, 2022) and to reduce experimenter demand effects (Stantcheva, 2022; Haaland et al., 2023). If respondents adapt their view that external circumstances are more decisive for educational success as a consequence of being informed about inequality in educational outcomes by parental background, we are interested in whether this also translates into higher private donation decisions to charities that aim at equalizing opportunities of students from different family backgrounds. In case our results show that treated respondents assign a higher importance to external circumstances as the decisive factor for educational success after being provided with the information about inequality of educational outcomes of students from different backgrounds, we also expect a positive treatment effect on respondents' donation decisions: especially if treated respondents perceive that the sources of inequality in students' chances lie beyond an individual's control, they are more inclined to donate (higher amounts) since they perceive the education system as less fair.

It is ex-ante unclear whether treatment effects on private donations will mirror its effects on public redistributive education spending. Examining redistributive education spending, we most strongly relate to the experimental literature that investigates how changing people's perceptions about the extent of inequality affects their preferences for governmental redistribution (see also Ciani et al. (2021) for a survey on the recent literature).[3] The results of the literature are rather mixed: while some studies find a positive effect on respondents' preferences for redistribution (Cruces et al., 2013), the common takeaway from the experimental studies is rather that while information usually leads to greater concern about inequality, it mostly fails to shift peoples' redistributional preferences neither towards policies aiming at equality of outcomes (e.g., Kuziemko et al., 2015; Hoy and Mager, 2021) nor towards many policies aiming at equality of opportunity (Alesina et al., 2018; Lergetporer et al., 2020). Respondents' policy preferences might differ from their donation decisions for multiple reasons which highlights the importance of examining both, respondents' donation decisions and preferences for redistributive education spending. In particular, while the process of increased spending is very clear for a donation to a charity, how public funds would increase if respondents stated a preference for higher targeted spending to help disadvantaged students is less certain. In particular, respondents might be unsure that increased education spending targeted at disadvantaged students by the government does not implicitly result in funds being diverted from more advantaged students. If treated respondents become more concerned about attendance levels, specifically of students from less advantaged backgrounds but also of students from more advantaged backgrounds to a lesser extent, they might not favor more public targeted spending. In contrast, the money from donations is added from the respondents' individual account to the charity and hence the origin and opportunity cost of funds is fully transparent to the respondent. Once informed about the unequal chances of students from less advantaged backgrounds, respondents might be especially willing to donate to charities that specifically target this group of students. Consequently, respondents are more willing to donate to a charity supporting students from disadvantaged backgrounds, about whom they seem to care more, while they are not more in favor of a change in the public-school spending funding formula.

It has been argued that the difference in treatment effects could also be due to a lack of trust in the government (Kuziemko et al., 2015) or doubts about the effectiveness of the policy proposal to achieve the goal of equality of opportunity. Imagine respondents hold the view that increased education spending is not the most effective means to promote equality of opportunity or that alternative policies may be more impactful. In that case, providing information about the inequality in educational outcomes of students from different backgrounds could alter respondents' private donation decisions, but not their preferences for governmental redistributive education spending as elicited in our specific policy question (see also Lergetporer et al. (2020) for a detailed discussion). A similar channel could be that

[3] The literature distinguishes between two types of policies, namely policies aiming at equality of outcomes, such as progressive taxation or minimum wages (e.g., Alesina and La Ferrara, 2005; Kuziemko et al., 2015) and policies aiming at equality of opportunity (e.g., Alesina et al., 2018; Lergetporer et al., 2020; Fehr et al., 2022b). Earlier work has provided survey respondents with information on their ranking in the national income distribution (Cruces et al., 2013; Kuziemko et al., 2015; Karadja et al., 2017; Hoy and Mager, 2021; Bublitz, 2022) or the global income distribution (Fehr et al., 2022a). Similarly, McCall et al. (2017) and Alesina et al. (2018) inform study participants about actual economic inequality in the U.S., and Lergetporer et al. (2020) and Fehr et al. (2022b) about the persistence of SES in Germany.

some respondents do not favor increases in governmental redistribution due to a lack of trust in the government (Kuziemko et al., 2015). Furthermore, how respondents perceive the role of circumstances or effort in achieving educational success could be influenced by their personal experience within the education system. For example, respondents who have achieved high educational degrees might attribute their success to their diligence and effort. Conversely, respondents who did not attain higher qualifications may lean towards attributing their failure to adverse conditions or external circumstances. Similarly, the effect of our information treatment on support for governmental redistributive education spending could differ by respondents' political ideology or partisanship. Left-leaning respondents are likely to express greater support of equal-opportunity policies undertaken by the government. In contrast, right-leaning respondents might change their perception on the role of circumstances in education and their preferences for private donations. However, their inclination might not extend towards additional government intervention (e.g., Alesina et al. (2018) on information about intergenerational mobility or Haaland et al. (2023) on information about racial gaps). We test these hypotheses with the data from the ifo Education Survey 2019 that we describe in section 4.4, and present the results after providing some information about the institutional background in Germany in the next section.

4.3 Institutional Background

In Germany, compulsory schooling usually starts at the age of six until the age of 18 (see Appendix Figure A4.1 for an overview of the German school system). The comprehensive primary school takes four years in the majority of states (some states have six years of primary education) and provides basic education in math, German, science, and social subjects. At the end of primary school, children are tracked into different school types. Some schools offer only basic and intermediate degrees that prepare for apprenticeship training or vocational education and usually last until grade 9 or 10, while other schools offer all tracks (Matthewes, 2021). Over time, the majority of German states has reformed tracking to increase the number of school types. *Gymnasium*, which we refer to as academic-track school throughout this paper, is the only secondary school type that exists in all states and that has remained unchanged by recent education reforms (Matthewes, 2021). This school type offers only an academic track which directly leads to the *Abitur*, the German university entrance qualification, after grade 12 or 13. Overall, academic-track school attendance is relatively common in Germany. 32 percent of 15-year-old children attended a *Gymnasium* in 2015 (own calculations based on PISA 2015 data).

Tracking decisions depend on parental preferences and the child's academic achievement at the end of primary school. Primary school teachers usually summarize their experience teaching a child and its grades in core subjects in a formal track recommendation. In 2015, in 12 of the 16 German states, this recommendation by the teacher was not binding, and it is at the parents' discretion to decide where to enroll their child for secondary school. In the remaining states, parents can only send their child to a higher track than recommended by the teacher if their child passes entry tests or performs well in trial lessons (see Grewenig (2021) for more details).[4]

Students are more likely to attend higher tracks if their parents have a high SES. Parental background even remains predictive of academic-track school attendance when test scores in math and reading are taken into account (own calculations based on PISA 2015 data). As a result, in international comparison, Germany has been repeatedly criticized for the fact that family background is a very strong predictor for students' educational performances compared to other countries. For instance, the German mean achievement gap in PISA 2015 science test scores that is associated with a one-unit increase on the PISA index of economic, social and cultural status amounted to 42 score points, the equivalent of more than one year of schooling, which lies above the average OECD performance gap of 38 score points (OECD, 2016). Similarly, while only 19 percent of 15-year-old children in the lowest 50 percent of families (in terms of their social background and family income) attend a *Gymnasium*, the respective share for children in the highest 50 percent of families amounts to 49 percent (own calculation based on data from PISA 2015, see Appendix A4.1 for details). The resulting gap of 30 percentage points is striking, not least because individuals with a university entrance qualification – which is typically obtained at a *Gymnasium* – do not only experience a large wage premium on gross earnings of around 42 to 44 percent (see Dodin et al. (2021) and Schmillen and Stüber (2014) for corresponding estimates), but also show lower risk of unemployment (Hausner et al., 2015) and higher life expectancy (Gärtner, 2002).

4.4 Data and Empirical Strategy

This section describes the data collection, the experimental design, sample characteristics, and the econometric model used for estimating the effects.

4.4.1 Data Collection and Sample

Our research is based on data from the ifo Education Survey 2019, a large opinion survey on education policy in Germany. Sampling and polling were conducted by Kantar Public, a renowned German survey company, in May 2019. Overall, the survey encompassed 37 questions related to education policy and respondents were also asked about a rich set of socio-demographic background characteristics at the end of the survey. Median completion

[4] In general, switching tracks or obtaining further qualifications after graduating from a lower track is possible, albeit rare. The yearly rate of changing school types is low, typically ranging from 1.3 percent (Baden-Württemberg) to 6.1 percent (Bremen). Among those students who changed the school type in 2010/11, only about 27 percent switched to a higher track school (Bellenberg, 2012). In addition, among all students who pursue secondary education after grade 10, more than 90 percent have attended a *Gymnasium* in the grades before (Statistisches Bundesamt, 2018). Therefore, the initial tracking decision after primary school is important in a child's educational career.

time was 30 minutes. Rates of item non-response are low, ranging between 0.0 percent and 0.2 percent for the questions used in this paper.

Respondents were sampled and surveyed via an online platform and answered the survey autonomously on their own digital devices. Respondents receive tokens by the online platform for their participation in the survey. In our survey, all respondents are incentivized with 75 tokens for survey completion. Subsequently, these tokens can be exchanged for items or gift vouchers of well-known online retailers. Thus, their exact value may differ between respondents depending on their personal preference for these items.[5]

To investigate belief updating and the persistence of potential information effects, respondents were also asked to participate in a follow-up survey roughly two weeks after completion of the main survey. The follow-up survey re-elicits respondents' belief about academic-track attendance rates of students from more and less advantaged backgrounds as well as other outcome variables but does not contain information about these attendance rates (identical for control and treatment group respondents). Overall, 80.2 percent of the original participants decided to take part in the follow-up survey. The median lag to the main survey was 15 days with a range from 7 to 40 days.

For our final analyses, we drop respondents who did not pass an attention check[6] posed half-way through the survey which leaves us with 2,094 respondents in the main survey and 1,671 respondents in the follow-up survey. As illustrated in Appendix Table A4.1, our sample is broadly representative of the German population in terms of gender, age, region, and household income. For instance, 79.6 percent of our respondents live in western Germany, compared to 80.3 percent in administrative data from the 2018 Microcensus.[7] Similarly, 53.1 (50.9) percent of respondents in our sample (in the administrative data) are female. Respondents in our sample are also reasonably close to the population in terms of their average age of 53.1 (50.9), a share of 41.3 (34.1) percent of sample respondents with university entrance degrees (*Abitur*), and a share of 43.8 (48.2 percent) of respondents above the respective me-

[5] Our compensation for survey participation corresponds to the standard rate that is offered by the polling firm. As an example, respondents may directly convert the 75 tokens into money, in which case they are worth about 0.75 Euro. This implies that the hourly wage equivalent of the compensation is relatively low, which already suggests that the collectable tokens may be (much) more valuable to the respondents than their pure monetary equivalent. Moreover, intrinsic motivation to state opinions or "gamification" – a phenomenon where respondents value tokens more than their monetary equivalent (e.g., Puleston, 2011; Keusch and Zhang, 2017) – might also increase survey participation.

[6] The wording of the attention check is as follows: "It sometimes happens that survey participants do not read individual questions accurately. To ensure that you read the questions accurately, we ask you to ignore the following question and enter the number twenty-two in the text field. [line-beak] The German states are also responsible for universities and colleges. What do you think, how many currently have tuition fees?" While none of the 16 German states currently has tuition fees, only respondents who answered 22 were left in the final sample.

[7] Research Data Centres of the Federal Statistical Office and the statistical offices of the Laender, Microcensus, census year 2018 (see also Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, 2018).

dian household income. Overall, our sample covers a broad and diverse range of individuals from the German population.

4.4.2 Experimental Design

Information Treatment

We conduct a survey experiment that informs respondents about the academic-track attendance rates of students from more and less advantaged backgrounds in Germany, i.e., the correlation between parental background and educational success of adolescents. Appendix Figure A4.2 provides an overview of the experimental design. In Germany, educational inequality between students from different socio-economic backgrounds is large in international comparison (e.g., OECD, 2020) and manifests early during children's educational careers. In fourth grade, children from more disadvantaged families show significantly lower skills in math, science, and reading (Stanat et al., 2017). This is particularly noteworthy as competencies achieved during primary school are decisive for students' transition to secondary school.

We study inequality in educational outcomes arising from the relationship between children's academic-track school attendance and their parents' SES. Since tracking into different school types occurs at a relatively young age for most children in Germany (usually between 10 and 12 years old), children's preferences for different school types are likely to be rather shortsighted, and their understanding of the consequences of education decisions is likely to be very limited. As a result, parental influence on the tracking decision is very high. At the same time, initial track choice creates a clear default educational outcome for students since changing school tracks at later ages is rare, academically difficult due to differences in teaching style and curriculum, and often comes with substantial social costs for students. Therefore, the decision whether or not a student attends an academic-track school after primary school is an important junction in the children's education journey that is highly dependent on parental initiative. Academic-track schools are the most popular school type where students obtain a university entrance qualification (see section 4.3 for more institutional details), which is positively associated with many favorable economic outcomes (e.g., Gärtner, 2002; Schmillen and Stüber, 2014; Hausner et al., 2015; Dodin et al., 2021). Differences in academic-track attendance thus capture a crucial aspect of equality of opportunity.

Our randomized information treatment informs respondents about the academic-track attendance rates of 15-year-old children in the lowest and highest 50 percent of family SES status. The treatment informs respondents that 49 percent of students from the more advantaged half of all families (in terms of their social background and family income) attend an academic-track school. Treated respondents also learn that, among students from the less advantaged half of all families, 19 percent attend an academic-track school. This results in an SES gap of 30 percentage points (see Appendix A4.1 for details about the calculation of the information treatment from PISA data). Along with the verbal statement about the academic-track attendance rates, respondents in the treatment group are also shown a graphical illustration of these attendance rates among students with different family backgrounds (see Appendix Figure A4.3 for details).

Eliciting Prior and Posterior Beliefs

To assess respondents' information status at baseline, we first elicit prior beliefs about the academic-track attendance rates of students from more and less advantaged backgrounds. Respondents are asked to report their best guesses for the shares of students from the more advantaged half and less advantaged half of all families (in terms of their social background and family income) who attend an academic-track school.[8] Given their guesses, we can also calculate the within-respondent estimate for the SES gap in attendance rates.

To shed further light on the belief updating process, we re-elicit respondents' beliefs about the academic-track attendance rates in the follow-up survey conducted two weeks after the main survey. For the control group, this captures any changes in their prior beliefs either due to priming or salience after participating in our survey or through other news or events in this period. For respondents in the treatment group, the follow-up survey allows us to study how respondents updated their beliefs due to receiving the information treatment in the earlier survey.

Eliciting the Perceived Role of Circumstances

We are interested in whether our factual information treatment on the academic-track school attendance rates changes respondents' perception of the role of students' effort and the role of circumstances in determining educational and labor-market success. We, therefore, ask respondents the following question: "Some say that success in life depends primarily on one's own effort. Others say that success in life depends primarily on external circumstances. In your opinion, what determines whether one achieves the following in life?" Respondents can then choose one of the following four answer categories "mainly own effort", "rather own effort", "rather external circumstances", or "mainly external circumstances". To analyze the extent to which respondents draw a connection between educational and economic success, we elicit these views for both the role of circumstances in achieving "a high educational degree" as well as "a high income".

Eliciting Private Donations

Next, we investigate whether information on academic-track attendance rates of students from more and less advantaged backgrounds changes respondents' desire to support students from disadvantaged backgrounds financially. Therefore, we implement a donation experiment, where respondents can choose to donate directly to charities that work to improve equality

[8] The corresponding belief elicitation question is posed to all respondents regardless of their treatment assignment and reads as follows: "Think of a comparison between children from the better and worse off half of all families (in terms of social background and family income). What do you think is the percentage of students from ... (i) the more advantaged half of all families who attend a *Gymnasium*?, (ii) the less advantaged half of all families who attend a *Gymnasium*?"

of opportunity in education. First, every respondent receives 80 tokens in addition to their regular compensation of 75 tokens for survey participation (see section 4.4.1 for details on the survey). Subsequently, respondents can decide to donate any amount between zero tokens or the full amount of 80 tokens to one or both of two charities that aim to help students from disadvantaged families.[9]

Given that donations directly reduce the monetary payout for the survey participants, they reflect revealed preferences for respondents' willingness to pay to support students from disadvantaged backgrounds. In addition, they are well suited to mitigate concerns of experimenter demand effects (Quidt et al., 2018; Mummolo and Peterson, 2019), as potential demand effects should be lower in tasks where real money is at stake (Haaland et al., 2023).

Eliciting Policy Preferences
We focus on respondents' policy preferences towards equality of opportunity, which is most directly relevant in the education context, and ask respondents whether they favor or oppose increased redistributive education spending by the government for children from less advantaged families to increase equality of opportunity.[10] Answers to this question are reported on a five-point Likert-scale ranging from "strongly favor" to "strongly oppose". The question also states that additional expenditures usually have to be financed through taxes.

4.4.3 Sample Balance

Appendix Table A4.2 presents results from a balancing test to check whether the randomization successfully balanced respondents' observable characteristics across the two groups of the main experiment. The first column shows the average characteristics in the control group. The subsequent columns present characteristics of the information treatment group together with the respective difference to the control group. With 31 comparisons, we would expect 1.55 to be significantly differently from zero at the five percent level. For our sample, two are significant at the five-percent level. Moreover, regressing treatment status simultaneously on all covariates yields a p-value for joint significance of 0.3. We thus conclude that random assignment worked as intended. We nevertheless include a large set of control variables in most regressions to increase the efficiency of our estimates.

[9] The selected charities are *Deutsches Kinderhilfswerk e.V.* and *Die Chancenstiftung*. Upon request, respondents could choose to learn more about these two charities by clicking on a link that displayed additional information (overall, 13.8 percent chose to learn more about the charities). The additional information about *Deutsches Kinderhilfswerk e.V.* states that the foundation is committed to a child-friendly Germany and that the donations are dedicated to the *"Chancengerechter Bildungsstart"* project, which, among others, provides children from low-income families with school materials. The additional information about *Die Chancenstiftung* states that the charity awards scholarships to children and young people from low-income families. The scholarship recipients usually receive professional tutoring.

[10] Measuring peoples' preferences for equality of opportunity policies by eliciting their view on education spending is a common approach (Alesina et al., 2018; Fehr et al., 2022b).

Next, Appendix Table A4.3 investigates whether participation in the follow-up survey is related to treatment assignment in the main survey. Regressing a dummy for follow-up-survey participation on the treatment indicator and covariates shows no evidence of differential attrition as a result of receiving the information treatment. The table further reveals that respondents who are older, those who tend to vote for parties outside of the mainstream, and respondents with a university entrance degree are more likely to participate in the follow-up survey. Reassuringly, among follow-up survey participants, respondents' observable characteristics are still well-balanced across treatment arms (see Appendix Table A4.4). Therefore, treatment-effect estimates of the information treatment on outcomes measured in the follow-up survey are still unbiased.

4.4.4 The Econometric Model

We estimate the effects of the information treatment on outcomes with the following regression model:

$$y_i = \alpha_0 + \alpha_1 T_i + \delta' X_i + \epsilon_i, \tag{4.1}$$

where y_i is the outcome variable of interest for respondent *i*, i.e., the perceived role of circumstances, private donations, and demand for redistributive education spending. T_i indicates whether respondent *i* received information on the relationship between academic-track attendance and parental background. X_i is a vector of control variables (see Table 4.1 notes for details), and ϵ_i is the error term. Since ϵ_i is uncorrelated with treatment status through randomization, the coefficient α_1 provides an unbiased estimate for the causal treatment effect of information provision. As the inclusion of covariates can increase the precision of estimates, we often show results both with and without covariates.

4.5 Results

This section describes the experimental results. We first examine whether information about the differential academic-track attendance rates by parental background affects respondents' perceived role of circumstances (section 4.5.1). We then analyze whether this effect also translates into respondents' donation decisions (section 4.5.2) and their preferences for redistributive education spending (section 4.5.3).

4.5.1 Information Provision and the Perceived Role of Circumstances

Figure 4.1 illustrates the treatment effect of providing information about the academic-track attendance rates of students from more and less advantaged students on respondents' perceived role of circumstances in achieving (i) a high educational degree (Panel A) and (ii) a high income (Panel B).

The information treatment has a large and significant effect on respondents' expressed view that educational attainment is the result of external circumstances rather than effort (see Figure 4.1, Panel A). In the uninformed control group, a baseline share of 17.3 percent states that a high educational degree (mainly or rather) depends on external circumstances, and 82.7 percent say that it is (mainly or rather) due to own effort. In the treatment group, where respondents are informed about the academic-track attendance rates of students from more (49 percent) and less advantaged families (19 percent), the share that attributes educational attainment to external circumstances largely and significantly increases by 12.1 percentage points to 29.4 percent. This finding is in line with what we expect if respondents care to a larger extent about attendance rates from less advantaged students.

The effects of information provision on the perceived role of circumstances in income inequality are similar but much smaller. Among respondents in the control group, 35.0 percent state that a high income is due to external circumstances rather than own effort (see Figure 4.1, Panel B). This share increases slightly to 38.8 percent, an increase which is nominally small and statistically insignificant at conventional levels.

Table 4.1 reports experimental results based on equation (4.1) as elicited on the four-point scale, with higher values indicating a stronger role of external circumstances.[11] Consistent with Figure 4.1, information on academic-track attendance rates of students from more and less advantaged backgrounds significantly shifts respondents' answers towards reporting a higher importance of external circumstances in determining high educational attainment (column 1). The effect amounts to 0.3 scale points on a scale from one to four which is an increase of 14.2 percent of the mean. Together with the finding of an increase in the share of those attributing educational success to external circumstances by 12.1 percentage points (69.9 percent), this is a rather sizeable effect. Column 3 confirms that the information treatment hardly affects respondents' view on the role of external circumstances in determining income. Reassuringly, including covariates does not qualitatively affect our results (columns 2 and 4).

Table 4.2 shows that significant effects of information provision persist in the follow-up survey two weeks after the information was provided.[12] It turns out that the initial treatment effect on the perceived role of circumstances for a high educational degree is very similar for the sample of respondents that participate in the follow-up survey, with an increase of 0.3 scale points (columns 1 and 2). The treatment effect of information provision on the same item in the

[11] Results are robust to the coding and choice of specification. While all models in this paper are estimated as linear probability models, (ordered) probit models yield qualitatively similar results (results available upon request). Similarly, conclusions drawn from Table 4.1 remain unchanged if we regress the treatment indicator on a binary indicator of perceived role of circumstances (analogous to Figure 4.1) or if a separate coefficient is estimated for each answer category (see Appendix Table A4.5).

[12] To analyze the effects of the information treatment on belief updating, we estimate the following model:

$$y_{it} = \beta_0 + \beta_1 T_i + \beta_2 T_i \text{ x Time surveyed}_t + \beta_3 \text{Time surveyed}_t + \delta' X_i + \epsilon_i, \tag{4.2}$$

where y_{it} is the outcome variable of interest for respondent *i* at time *t* with $t \in$ {main survey, follow-up survey}.

follow-up survey is reduced by about two-thirds but is still positively significant. The smaller size of the coefficient in the follow-up survey would be expected and could be consistent with imperfect recall or information compliance. Nevertheless, the positive effect implies that respondents in the treatment group change their perception of the role of circumstances in education not only momentarily while the information on academic-track attendance rates by parental SES is available on the screen but over an extended time period. While both salience effects and belief updating can drive immediate effects of information provision, this persistence implies that respondents are indeed able to understand and remember the provided information. Therefore, the treatment effect is rather unlikely to stem only from salience or experimenter-demand (Haaland et al., 2023) while parts of it stem from information updating.

In sum, providing information on the extent of inequality in academic-track attendance rates by parental background has a large and positive effect on the share of respondents who view a high educational degree as the result of external circumstances rather than effort, which persists in a follow-up survey two weeks later. At the same time, it does not affect respondents' perception of the role of circumstances in determining income inequality, suggesting that respondents do not infer a strong link between circumstance-induced differences in educational outcomes and income differences. This runs counter to a large literature documenting the importance of educational outcomes for future earnings (e.g., Schmillen and Stüber, 2014; Dodin et al., 2021). An alternative interpretation would be in line with a literature on equality of opportunities, which conjectures that adults are responsible for their own education and career decisions, while children might not be (Roemer, 2004). In this case, respondents might believe that earnings inequality is predominantly affecting adults, whom they see as responsible for their labor-market outcomes regardless of their parental circumstances.

4.5.2 Information Provision and Private Donations

In this section, we report estimates showing that information on inequality in educational outcomes of students from different parental background increases donations to charities supporting students from disadvantaged backgrounds for respondents in the treatment group. At baseline, most respondents (66.2 percent) decide to donate a positive amount to charities.

Table 4.3 regresses donations on the treatment indicator based on equation (4.1). Information on the difference in academic-track attendance rates by parental background significantly increases the share of respondents who decide to donate any positive amount by 9.3 percentage points (see column 1). Similarly, the average amount of donated tokens increases significantly by 3.3 tokens in the treatment group, compared to 37.5 tokens at baseline (column 2). The share of respondents who decides to donate the full amount of 80 tokens (30.3 percent) does not change in the treatment group (column 3). We also find an increase in the share of respondents who donate more than the control group median (column 4), suggesting that our information treatment does not only positively affect very small donations.[13]

Figure 4.2 additionally shows the treatment effect on donations. Information about the extent of inequality in academic-track attendance rates by parental background particularly decreases small donations, i.e., those up to an amount of 10 tokens. In contrast, the treatment positively affects donations larger than 10, except those between 40 and 50 tokens and between 70 and 80 tokens.

4.5.3 Information Provision and Demand for Redistributive Education Spending

In this section, we show that positive treatment effects of providing information on academic-track attendance rates of students from more and less advantaged backgrounds do not translate into increased demand for redistributive education spending by the government. Table 4.4 regresses respondents' policy preferences about redistributive education spending on the treatment indicator based on equation (4.1). Depending on the exact specification, effect sizes vary between -1.1 percentage points (support in column 1) and 1.5 percentage points (opposition in column 2) and are neither statistically nor economically significant. Further exploiting variation by measuring preferences on the continuous five-point measure of support for more redistributive education spending shows very similar results. In this specification, the information treatment decreases demand for redistributive education spending by -0.005 points on the Likert scale, which is not statistically significant at conventional levels (column 3).

Descriptively, we find a very high baseline support towards increased governmental spending for children from less advantaged families. In the control group, 75.1 percent of respondents (strongly) favor increased education spending to foster equality of opportunity. Only a small minority of 12.6 percent opposes higher spending for children from disadvantaged backgrounds.

Despite the considerable support among respondents for education spending aimed to foster equality of opportunity, and the fact that information on academic-track attendance rates of students from more or less advantaged backgrounds changes the perception of the role of circumstances in education and private donation behavior, we do not find a strong effect of the information treatment on policy preferences in the studied sample. This is in line with previous literature showing that policy preferences for redistributive reforms can be unresponsive even if perceptions of inequality change (e.g., Alesina et al., 2018). Our findings in this section are also in line with the literature that finds that policy preferences in the general

[13] When making their donation decision, respondents had the opportunity to distribute their donations between two charities (see section 4.4.2 for details). In the control group, the majority of those who decide to donate a positive amount split their donations equally between both charities (63.7 percent). 28.6 percent allocate the full amount of their donations to *Deutsches Kinderhilfswerk e.V.* The remaining 7.6 percent allocate the full amount to *Die Chancenstiftung*. While our information treatment significantly increases average donations, the allocation of donations between the charities remains largely unaffected (results available upon request).

population are less malleable to information provision than other preferences. For example, Luttmer and Singhal (2011) suggest that preferences for governmental redistribution have an important cultural component that is rather stable over time. Other studies argue that even if respondents update their factual beliefs, it remains unclear whether people use these facts in forming political opinions (Gaines et al., 2007; Khanna and Sood, 2018; Zhang, 2022).

4.6 Mechanisms

In this section, we discuss additional evidence on belief updating (section 4.6.1), the role of academic achievement of students for respondents' preferences (section 4.6.2), and heterogeneities that might help characterize the parts of the population that are susceptible to information (section 4.6.3).

4.6.1 Evidence on Misperceptions and Belief Updating

First, we examine to what extent information effects are related to previous misperceptions and what part of the information treatment drives respondents' answering behavior. We first show descriptive evidence on respondents' prior beliefs for the share of children from different parental backgrounds that attend an academic-track school before information provision. We then provide experimental evidence on the effect of information provision on posterior beliefs from the follow-up survey about two weeks later (see section 4.4.2).

Overall, respondents severely misperceive both the levels and the SES gap of academic-track attendance in Germany. On average, respondents guess that 71 percent of students (accurate value: 49 percent) from a more advantaged parental background attend the academic-track school (see Appendix Figure A4.4 for the full distribution of guesses). At the same time, they also guess that 30 percent of students (accurate value: 19 percent) from a less advantaged parental background attend an academic-track school. These beliefs result in a misperception of the SES gap in academic-track attendance, which respondents expect to amount to 41 percentage points on average (accurate value: 30 percentage points). This would imply that the average treated respondent receives a downward information shock related to lower levels of attendance of academic-track schools compared to their prior belief (in case they favor high educational outcomes for all students) but a positive information shock related to the smaller SES gap between groups (in case they value equality of opportunity). Our finding that information provision leads to an increase in the perception that circumstances are important in achieving a high educational outcome (see section 4.5.1) therefore suggests that the average respondent is more concerned with the levels of academic-track attendance of students for less advantaged backgrounds than with the size of the difference between groups.

In line with this interpretation, we find the strongest evidence of belief updating in the follow-up survey for the guesses of academic-track attendance rates of students from less advantaged backgrounds. The follow-up survey re-elicits respondents' beliefs about the share of students from more and less advantaged families in the same way as in the main survey but does not include any reminder of information provision (see section 4.4.2). Table 4.5 regresses posterior beliefs on the treatment indicator from the main survey based on equation (4.1). It shows that information provision persistently improves beliefs about academic-track attendance rates. Respondents in the control group still estimate similar shares of students from less (30.3 percent) and more (69.0 percent) advantaged backgrounds to attend the academic-track schools as the control group in the main survey. Treated respondents' beliefs about both these shares decrease, which is in line with lower academic-track school attendance rates for both groups. Respondents' beliefs about the academic-track attendance rates of students from less advantaged backgrounds decrease by 1.9 percentage points or 6.4 percent, while beliefs for students from more advantaged backgrounds decrease by 0.8 percentage points or 1.1 percent. While effect sizes go in the expected direction for both guesses, the effect of the information treatment only reaches statistical significance for the guess of the share of students from less advantaged backgrounds. Taken together, these findings suggest that respondents care most about the number of students going to academic-track schools, rather than differences between groups, and are most concerned about students from less advantaged backgrounds. This result is in line with the finding that treated respondents are more willing to donate to help disadvantaged students. It also supports the interpretation that preferences for targeted public spending do not increase in the treatment group as respondents are unwilling to redirect education funding from students from more advantaged backgrounds, which similarly perform below expectations.

In the context of redistributive education spending, respondents may be uncertain about whether increased governmental education spending on the education for disadvantaged students might involve a reduction in resources allocated to more advantaged students. As respondents are exposed to information which increases concerns about the level of attendance rates – particularly among students from less advantaged backgrounds, but to a lesser degree also among students from more advantaged backgrounds – their tendency to support increased redistributive education spending may be less stable. This stands in contrast to the transparency exhibited in charitable donations, where funds are distinctly sourced from the respondents' individual accounts and seamlessly channeled to the designated cause, leaving no ambiguity about the funds' origins or potential trade-offs. Upon learning about the inequality in educational outcomes by parental background, respondents may display a higher willingness to contribute to charities that address the needs of less advantaged students. Nevertheless, this increased disposition towards charitable giving does not necessarily translate to a parallel support of altering the existing funding allocation formula.

4.6.2 Role of Academic Achievement

Second, we examine how respondents view the role of academic achievement in determining academic-track school attendance rates of students from different backgrounds. In a meritocratic school system, a common expectation is that educational success, including attendance of an academic-track school, should depend on ability and academic achievement rather than on circumstances or effort directly. Therefore, respondents could expect that more advantaged children are more likely to attend academic-track schools because they are better in school, and that the resulting SES gap in academic-track attendance purely reflects higher academic achievement of students from more advantaged parental backgrounds. Alternatively, respondents might perceive differences in attendance rates to reflect parental preferences for vocational education (Lergetporer et al., 2021) or behavioral, informational or institutional barriers.

In order to understand respondents' perceptions of these two mechanisms, we extend our experimental framework by providing a third experimental group with an extended information treatment. Respondents in this group receive the same information as the main treatment group (see section 4.4.2), but in addition receive information on the residual SES gap in academic-track school attendance after controlling for students' academic achievement, which is equal to 16 percentage points (see Appendix A4.1). This extended treatment allows us to investigate to what extent perceived differences in academic achievement drive any treatment effects of information provision on our outcome measures. If respondents assume that the SES gap in academic-track attendance rates is primarily driven by differences in academic achievement (i.e., if they underestimate the residual SES gap), receiving information on the residual SES gap could prompt respondents to think that students from different backgrounds do not differ as much as expected in terms of their academic achievement. In that case, the extended treatment likely yields larger treatment effects than the shorter treatment as respondents receive a downward shock on the perceived importance of academic achievement for academic-track attendance rates. Conversely, if respondents assume differences in academic achievement between groups to be small, the information that the difference in attendance rates conditional on achievement is 16 percentage points might lead respondents to conclude that a larger proportion of the gap than anticipated can be attributed to differences in academic achievement which might limit their view that circumstances are the most decisive factor and treatment effects might be smaller.

As it turns out, treatment effects on the perceived role of circumstances of the extended information treatment are smaller than those of the shorter treatment that does not include information on the conditional gap in academic-track attendance rates (see Appendix Table A4.6). This suggests that respondents underestimate differences in academic achievement for students from different parental backgrounds if this information is not included. The difference between the two treatment effects on the perceived role of circumstances for a high educational degree is statistically significant (Appendix Table A4.6, column 1).

Interestingly, treatment effects of the second treatment on the average amount of donated tokens (2.9 tokens) are not significantly different between the two treatments, even though the point estimate is slightly lower (column 3).[14] This suggests that respondents' concern for the levels of academic-track school attendance of students from less advantaged backgrounds (see section 4.6.1) is not diminished by information on differences in academic achievement between groups. This would be in line with the interpretation that respondents do not reduce their support for targeted spending benefiting students from less advantaged backgrounds if they receive information that academic achievement differences are larger than expected.

Next, we study treatment effects on the stated importance of different aspects related to the transition from primary to secondary schools. We ask respondents to rate how important the following five aspects are to determine whether a student transitions to an academic-track school on a five-point scale from "very important" to "very unimportant": (i) "educational attainment of parents", (ii) "financial situation of parents", (iii) "effort and diligence of students", (iv) "talent of students", and (v) "preferences of students and parents". We discuss effects on these factors in turn.

In the control group, 62.5 percent of respondents state that the educational background of parents is very or rather important for the transition to the highest academic-track schools, and 55.7 percent express the same sentiment for the financial situation of parents. Regressing the importance of the different aspects on the treatment indicator reveals that information about the extent of inequality in academic-track attendance rates by parental background increases the importance that respondents assign to the educational background and financial situation of parents (see Table 4.6).[15] We find effect sizes of 0.1 standard deviations for both aspects. This finding is consistent with respondents' increased perception of external circumstances beyond an individual's control as the decisive factor for educational success or outcomes.

For the third item, the importance of effort, a very high share of respondents in the control group of 93.3 percent thinks that effort and diligence of students is important for the transition from primary to secondary school. Interestingly, respondents' stated importance of students' effort and diligence is hardly affected by the information treatment, suggesting that respondents do not revise their view on the importance of effort when they update their beliefs about the importance of individual circumstances. Thus, while respondents were less likely to believe that effort rather than external circumstances was the decisive factor for determining educational success, this result suggests that they might continue to regard it as a necessary precondition. Only half of the control group respondents consider the preferences of students and parents as important for the transition from primary to secondary schools while 90.7 percent think that students' talent is important. We do not find effects of information provision on the perceived importance of students' talent or students' and parents' preferences for an academic-track school, suggesting that respondents do not update their perceptions of parental preferences when receiving the information treatment.

[14] In line with the findings on the main treatment effect, we do not find a treatment effect on preferences for redistribution (column 4).
[15] For the regressions, we z-standardize the five-point scale outcomes.

Taken together, results on the perceived importance of different aspects suggest that treated respondents especially change their view towards the importance of external circumstances, i.e., parental education and financial situation, for determining students' track choice at the transition from primary to secondary school. Thus, our finding that a larger share of respondents in the treatment group states that external circumstances rather than effort determine educational outcomes seems to be driven by an increase in the perceived role of circumstances, rather than a decrease in the perceived importance of effort. This observation also corroborates the interpretation that donation behavior is driven by an increased perception of respondents that there exist more institutional barriers to academic-track school attendance than expected, that cannot be overcome with students' effort alone.

4.6.3 Policy Effectiveness and Heterogeneity of Treatment Effects

Lastly, we investigate whether our patterns of results, in particular the absence of treatment effects on support for redistributive education spending, is driven by offsetting effects in particular subgroups or concerns regarding the political process.[16]

First, we explore whether doubts about policy effectiveness can explain our limited treatment effects on policy preferences. Suppose respondents believe that targeted support by charities can support less advantaged students, while public redistributive education spending is not suitable to foster equality of opportunity. In that case, information provision about the extent of inequality in academic-track attendance rates by parental background may induce respondents to increase private donations, but may not change their preferences for redistributive education spending by the government as elicited in our specific policy question.

To test this hypothesis, we asked a subset of respondents whether, in their opinion, a variety of potential policy interventions are suitable to foster equality of opportunity. Appendix Figure A4.5 reveals that the vast majority of respondents (84.2 percent) states that increasing governmental expenditure to schools mostly serving children from a disadvantaged family background is very or rather suitable to decrease educational inequality in Germany (Panel A). In fact, this share is among the highest when compared to other policy proposals frequently discussed in the context of reducing educational inequality in the German public debate. Therefore, we conclude that a perceived ineffectiveness of public targeted spending to achieve the policy goal is an unlikely driver of our results.

Similarly, treatment effects could be limited if respondents expect increased education spending to jeopardize other important goals of education policy. In particular, respondents may

[16] In the pre-registration for this experiment, we committed to performing heterogeneity by prior beliefs, respondents' own educational attainment, and respondents' trust in government. This section additionally reports explorative results on political ideology. The analysis of prior beliefs shows that differences in treatment effects between over-estimators, under-estimators, and those whose guesses are roughly correct are not statistically significant, although sizes of point estimates are non-negligible (results available on request).

perceive a trade-off between increasing equality and increasing efficiency of the education system. We do not find evidence of this belief in our sample: as Appendix Figure A4.5 reveals, the vast majority (80.8 percent) states that increased governmental spending for schools is also very suitable to increase the performance of the German education system (Panel B).

Next, we explore the role of respondents' socio-demographic characteristics. Respondents' perception of the role of circumstances and effort in achieving educational success might be partly based on their personal experience in the education system. For example, respondents who have high educational degrees might believe their success was due to their diligence and effort. In contrast, respondents who did not obtain higher qualifications might be more likely to attribute their failure to adverse conditions or external circumstances. Therefore, the information might lead to asymmetric updating for respondents with different education degrees. Appendix Table A4.7 reports heterogeneity of the treatment effects by whether respondents obtained a university entrance qualification (*Abitur*, which is usually obtained at academic-track schools, see section 4.3). It turns out that for private donations, the information effect is driven predominantly by respondents who do not have a university entrance qualification. Treatment effects on the perception of the role of circumstances and demand for redistributive education spending do not differ between groups. These findings are in line with the interpretation that donation behavior is driven by the concern for the academic school-track attendance of students from less advantaged backgrounds, and does not suggest that information updating differs between respondents with different education backgrounds.

Also, the null effect of our information treatment on support for redistributive education spending by the government could mask important heterogeneities by respondents' political ideology or partisanship. Left-wing respondents may be more supportive of equal-opportunity policies undertaken by the government. In contrast, right-wing respondents may indeed change their views on the role of circumstances in education and preferences for private donations but not favor additional governmental intervention (e.g., Alesina et al. (2018) on information about intergenerational mobility or Haaland and Roth (2023) on information about racial gaps). We test this hypothesis using data on respondents' long-term party attachment and distinguish between the following three subgroups[17]: (i) left-leaning partisans, i.e., respondents who report that they support SPD, LINKE, or GRÜNE, (ii) right-leaning partisans, i.e., respondents who report that they support CDU/CSU, FDP, or AfD and (iii) non-partisans, i.e., respondents who report that they have no particular long-term party attachment. Appendix Table A4.8 reports results on the perceived role of circumstances, private donations, and demand for redistributive education spending. Columns 1, 4, and 7 report information effects for left-leaning respondents, columns 2, 5, and 8 for right-leaning respondents and the remaining columns for respondents with no particular party attachment. The information treatment effect on respondents' view that mainly external circumstances rather than effort determine educational success is stronger for left-leaning respondents than for right-leaning respondents.

[17] We focus on long-term party attachment because it reflects fundamental political values instead of short-term considerations guiding intended voting behavior.

Similarly, treatment effects on private donations are larger, albeit not statistically significant, for respondents with any party affiliation compared to those who are non-partisans. However, treatment effects on preferences for governmental redistributive education spending are not significantly different between respondents with different political ideologies, even though left-leaning respondents report more demand for increases in redistributive education spending at baseline than right-leaning respondents and non-partisans (see columns 7, 8 and 9). Overall, differences by political ideology appear too small to explain our overall pattern of results.

A similar channel could be that some respondents do not favor increases in governmental redistribution due to a lack of trust in the government (Kuziemko et al., 2015). Indeed, 67.9 percent of respondents report that they have little or no trust in the German government. As it turns out, we find no evidence of significant treatment differences for the perception of the role of circumstances and demand for redistributive education spending (see Appendix Table A4.9). The coefficient for the treatment dummy on private donations is substantially larger for respondents reporting high trust in the German government but power of the subgroup analysis is limited (see column 3, p-value = 0.09).

Overall, we find some evidence of heterogeneity on private donations, especially by respondents' own educational background, but no indication of heterogeneous treatment effects on demand for redistributive education spending. In order to further decrease the likelihood of undetected heterogeneous effects in subgroups that could change the interpretation of our main findings, we implement a data-driven machine learning approach, the Causal Forest algorithm (Breiman, 2001; Bertrand et al., 2017; Wager and Athey, 2018; Athey et al., 2019). This approach also allows us to capture more complex, high-dimensional combinations of covariates that might be missed otherwise (see Appendix A4.2 for more details on the method). Appendix Figure A4.6 visualizes the distribution of the predicted Conditional Average Treatment Effects (CATEs) on demand for redistributive education spending. Overall, the predicted treatment effects are not indicative of treatment heterogeneity. The majority of the predicted CATEs are between -0.10 and 0.10, which is relatively small compared to the scale from one to five. This impression is confirmed when dividing the sample into four subgroups according to the size of their predicted CATE and calculating the average treatment effect within these four groups (Appendix Figure A4.7). Appendix Table A4.10 shows the differences between the four groups in respondents' characteristics. We see differences between respondents with the lowest predicted CATE and respondents with the highest predicted CATE in most dimensions, although the magnitudes are often not economically important. Finally, Appendix Table A4.11 shows the ranking of the covariates in terms of the variable importance. The variable importance captures the relative frequency with which a forest splits on the covariates across all grown trees (Farbmacher et al., 2021). Interestingly, whether respondents hold a university entrance qualification is one of the most important variables, even though effect heterogeneities were confirmed to be small in Appendix Table A4.7.

In sum, we observe limited heterogeneity of treatment effects for the perception of the role of circumstances and demand for redistributive education spending. Treatment effects on private donations seem to show some heterogeneity by respondents' own educational background. We also rule out several explanations for the muted treatment effect on respondents' demand for redistributive education spending. Lack of trust in government, partisan biases, doubts about policy effectiveness and own educational attainment cannot explain the difference in effects between private donations and policy preferences, leaving differences between the transparency of the spending process for donations and education spending as a likely explanation.

4.7 Conclusion

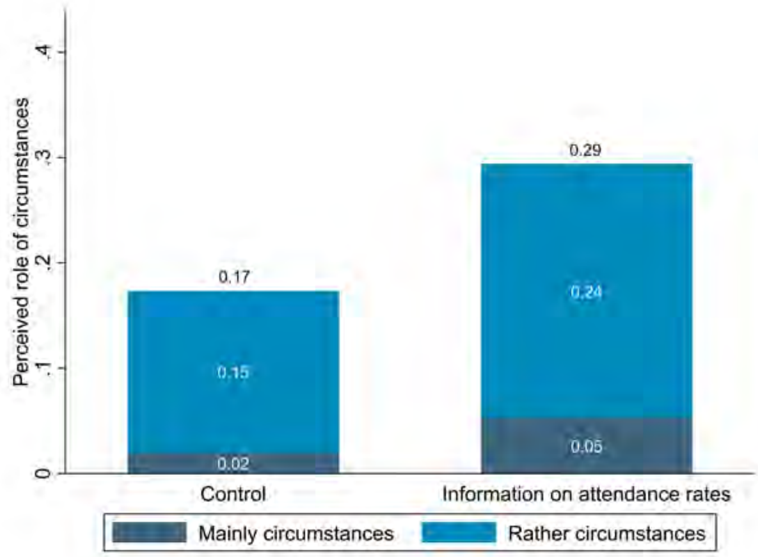
Educational inequality is a major concern of policymakers around the world and could play a crucial role in determining demand for targeted spending in favor of less advantaged students. By conducting a large-scale experiment, we study how information about academic-track school attendance rates by parental background in Germany affects individuals' perception of the role of circumstances, their preferences for private donations, and their demand for governmental redistributive education spending. The provided information consists of three pieces: the absolute academic-track attendance rates between students from more and less advantaged backgrounds as well as the difference in these attendance rates of those two groups. We find that most respondents think that educational success is determined by effort rather than external circumstances. We then show that information about academic-track attendance rates of more and less advantaged students in Germany affects these views: when information on attendance rates is provided, the share of respondents who believe that success in education is determined by circumstances significantly increases. Similarly, respondents in the treatment group increase donations to charities that provide materials and resources to students from less advantaged backgrounds, while demand for public targeted spending remains unchanged.

We document that respondents underestimate the academic-track school attendance rates of students from both less and more advantaged backgrounds, which would suggest positive effects on support for targeted spending, while they are overly pessimistic regarding the size of the difference between groups, which might mean that information on the true gap reduces support for less advantaged students. Our finding of positive treatment effects on donation behavior thus suggests that respondents are predominantly concerned with the levels of academic-track attendance in Germany, rather than with differences between different groups of students. We also document that information retention is highest for information on the academic-track attendance rates of less advantaged students, which suggests that this is the piece of information respondents might find most relevant. Differences in treatment effects for donation behavior, i.e., private targeted spending, and public redistributive spending are in line with the interpretation that respondents support targeted spending if this funding is not redirected for more advantaged students or other areas of the education system. While redistribution in the donation setting takes places between respondents themselves and charities that support students from less advantaged backgrounds, the opportunity costs of funds redirected through public spending formulas are less clear. Since respondents also *overestimate* the educational achievements of students from more advantaged backgrounds, this suggests they might be reluctant to support public redistributive policies upon learning both groups' academic-track attendance rates.

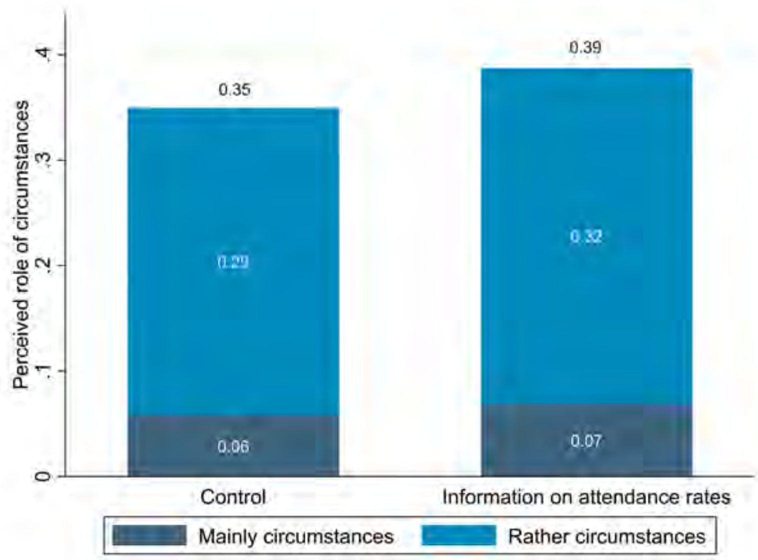
Our results speak to the political economy of education finance in Germany. Individuals show a high willingness to support students from less-advantaged backgrounds through private or public targeted spending, and increase their private support when they receive information on educational outcomes of different groups of students. Whether information on the education prospects of disadvantaged students can also create a politically feasible pathway to more targeted education spending on a large scale remains an open question for further research. More research is needed to explore whether transparency regarding the source and opportunity costs of governmental spending is the missing link between respondents' concerns and support for targeted public spending.

Figures and Tables

Figure 4.1: Effect of Information Treatment on Perceived Role of Circumstances



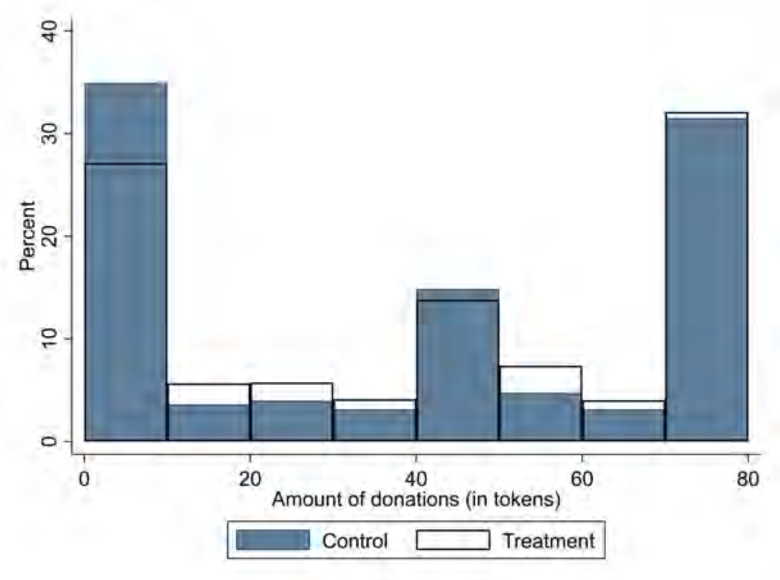
Panel A: Perceived Role of Circumstances (High Educational Degree)



Panel B: Perceived Role of Circumstances (High Income)

Notes: Responses to the question "Some say that success in life depends primarily on one's own effort. Others say that success in life depends primarily on external circumstances. In your opinion, what determines whether one achieves the following in life? A high educational degree (Panel A); A high income (Panel B)". Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Data source: ifo Education Survey 2019.

Figure 4.2: Distribution of Private Donations across Experimental Groups



Notes: Distribution of respondents' private donations to charities that aim at supporting students from less advantaged backgrounds to foster equality of opportunity (divided in eight bins) by treatment and control group. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Data source: ifo Education Survey 2019.

Table 4.1: Effect of Information Treatment on Perceived Role of Circumstances

	Perceived role of circumstances (high educational degree)		Perceived role of circumstances (high income)	
	(1)	(2)	(3)	(4)
Information on attendance rates	0.255***	0.257***	0.048	0.048
	(0.036)	(0.035)	(0.038)	(0.037)
Covariates	No	Yes	No	Yes
Control mean	1.802	1.802	2.181	2.181
Observations	2,094	2,094	2,093	2,093
R-squared	0.024	0.073	0.001	0.042

Notes: OLS regressions. Dependent variables: (1) – (2) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances ("mainly external circumstances", "rather external circumstances", "rather own effort", "mainly own effort"); (3) – (4) Perception that circumstances, not effort, are decisive for high income on a four-point scale, with higher values indicating more importance of circumstances ("mainly external circumstances", "rather external circumstances", "rather own effort", "mainly own effort"). Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable in the control group. Covariates include: age, female, born in Germany, West Germany, living in large city, risk, patience, parents with university education, income, current employment status, middle school degree, high school degree, partner living in household, parental status, work in education sector and imputation dummies. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.2: Persistence of Information Treatment Effect on Perceived Role of Circumstances

	Perceived role of circumstances (high educational degree)		Perceived role of circumstances (high income)	
	(1)	(2)	(3)	(4)
Information on attendance rates	0.262***	0.256***	0.054	0.052
	(0.040)	(0.039)	(0.042)	(0.042)
Information on attendance rates x follow-up	-0.183***	-0.183***	-0.036	-0.036
	(0.040)	(0.040)	(0.043)	(0.043)
Follow-up	-0.017	-0.017	-0.049*	-0.049*
	(0.028)	(0.029)	(0.030)	(0.030)
Information on attendance rates in follow-up	0.079**	0.074**	0.018	0.016
	(0.036)	(0.035)	(0.041)	(0.040)
Covariates	No	Yes	No	Yes
Observations	1,671	1,671	1,671	1,671
R-squared	0.020	0.063	0.002	0.042

Notes: OLS regressions. See equation (4.2). Dependent variables: (1) – (2) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances; (3) – (4) Perception that circumstances, not effort, are decisive for high income on a four-point scale, with higher values indicating more importance of circumstances. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Sample: respondents who participated in the follow-up survey. Robust standard errors in parentheses, adjusted for clustering at the respondent level. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.3: Effect of Information Treatment on Private Donations

	Any positive donation	Average donations	Full donation	Donation above median
	(1)	(2)	(3)	(4)
Information on attendance rates	0.093***	3.267**	0.004	0.046**
	(0.020)	(1.401)	(0.020)	(0.021)
Covariates	Yes	Yes	Yes	Yes
Control mean	0.662	37.499	0.303	0.396
Observations	2,093	2,093	2,093	2,093
R-squared	0.056	0.061	0.046	0.041

Notes: OLS regressions. Dependent variables: (1) Dummy variable coded one if amount of donations is positive; (2) Amount of donations stated by respondents (in tokens); (3) Dummy variable coded one if amount of donations is 80 (maximum possible donation); (4) Dummy coded one if amount of donation is above the control group median donation. Randomized experimental treatment “information on attendance rates”: respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.4: Effect of Information Treatment on Demand for Redistributive Education Spending

	Support for increased redistributive educ. spending	Opposition to increased redistributive educ. spending	Demand for redistributive educ. spending
	(1)	(2)	(3)
Information on attendance rates	-0.011	0.015	-0.005
	(0.019)	(0.015)	(0.043)
Covariates	Yes	Yes	Yes
Control mean	0.751	0.126	3.823
Observations	2,094	2,094	2,094
R-squared	0.034	0.023	0.040

Notes: OLS regressions. Dependent variables: (1) Dummy variable coded one if respondent is mainly/rather in favor of increased redistributive education spending; (2) Dummy variable coded one if respondent is mainly/rather against increased redistributive education spending; (3) Demand for redistributive education spending on five-point scale, with higher values indicating more demand for redistribution. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5: Effect of Information Treatment on Posterior Beliefs Elicited in the Follow-Up Survey

	Belief: academic-track attendance (SES gap)		Belief: academic-track attendance (high SES)		Belief: academic-track attendance (low SES)	
	(1)	(2)	(3)	(4)	(5)	(6)
Information on attendance rates	0.942	1.154	-0.869	-0.784	-1.811***	-1.938***
	(1.033)	(1.032)	(0.813)	(0.808)	(0.668)	(0.671)
Covariates	No	Yes	No	Yes	No	Yes
Control mean	38.689	38.689	68.957	68.957	30.268	30.268
Observations	1,671	1,671	1,671	1,671	1,671	1,671
R-squared	0.000	0.040	0.001	0.042	0.004	0.028

Notes: OLS regressions. Dependent variables: (1) – (6) Respondents' stated posterior belief as indicated in the table header. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Sample: respondents in the follow-up survey. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.6: Effect of Information Treatment on Aspects Important for Academic School Attendance

	Parental education	Financial situation	Effort	Talent	Preferences
	(1)	(2)	(3)	(4)	(5)
Information on attendance rates	0.134***	0.124***	-0.015	-0.017	0.070
	(0.043)	(0.043)	(0.040)	(0.040)	(0.044)
Covariates	Yes	Yes	Yes	Yes	Yes
Control importance	0.625	0.557	0.933	0.907	0.500
Observations	2,094	2,094	2,094	2,094	2,094
R-squared	0.047	0.037	0.050	0.075	0.015

Notes: OLS regressions. Dependent variables: Respondents' stated importance that the following aspects are important for transition to *Gymnasium* elicited on five-point scale, 1 = not important at all, 5 = very important, standardized mean zero, standard deviation one; (1) Parental education; (2) Financial situation; (3) Students' effort; (4) Students' talent; (5) Students' and parents' preferences. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control importance: share of those who state that respective aspect is very/rather important in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Appendix A4.1 Derivation of Information Treatments on Educational Inequality

Our randomized information treatment informs respondents about the gaps in academic-track school (*Gymnasium*) attendance rates of 15-year-old children in the lowest and highest 50 percent of family SES status. The main treatment informs participants that 49 percent of students from the more advantaged half of all families (in terms of their social background and family income) attend a *Gymnasium* while only 19 percent of students from the less advantaged half of all families do so. This yields an unconditional SES gap of 30 percentage points.

The information on the SES gap provided in the treatment could, for instance, reflect the fact that low SES students perform worse in school and are therefore less likely to attend the *Gymnasium*. Alternatively, it could reflect SES differences in behavioral barriers (e.g., institutional knowledge of parents) that are unrelated to student achievement.

The treatment uses the connection between children's school attendance and their parents' socio-economic status as a measure for educational inequality. The focus on academic-track school attendance captures an important dimension of equality of opportunity since *Gymnasium* attendance is a crucial step towards obtaining the university entrance degree and, thus, also important for later life income (e.g., Dodin et al., 2021).

To calculate the gap in academic school attendance rates, we use data from the Programme for International Student Assessment (PISA) conducted by the OECD in 2015. For the SES split, we use the PISA index of economic, social, and cultural status (ESCS), a composite measure of home possessions including books at home, the highest parental occupation, and the highest parental education. We first rank German children according to their points in this index and then perform a median split of students. We calculate that 19 percent children with an SES index score below the median and 49 percent of children with an index score above the median attend a *Gymnasium*.

As mentioned in section 4.6.2, we also provide a third experimental group with an additional information treatment. Respondents in this group receive the same information as the main treatment group (see section 4.4.2), as well as information on the residual SES gap in academic-track school attendance when controlling for students' academic achievement. Thus, this treatment further informs respondents that if we compare only students who are equally good in math and reading, the SES gap amounts to 16 percentage. To obtain the conditional SES gap, we regress *Gymnasium* attendance on a dummy indicating whether the student is above or below the median of SES index alongside controls for the students' math and reading test scores in PISA. We then use the coefficient on the SES indicator to report a conditional gap of 16 percentage points.

Measuring educational inequality as socio-economic differences in *Gymnasium* attendance has a major advantage. In contrast to achievement measures, e.g., PISA test scores as used, for instance, by Lergetporer et al. (2020), *Gymnasium* attendance rates are easily interpretable for the general population. In the public debate, differences in academic school attendance rates are frequently used by the media to report on the extent of educational inequality. For instance, the newspaper ZEIT has several reports on the so-called *Bildungstrichter* ("education hopper") with an essential component of this hopper constituting the difference in *Gymnasium* attendance between high- and low-SES students (see, e.g., Spiewak, 2018).

Appendix A4.2 Causal Forest Algorithm

A4.2.1 Theory

We use the Causal Forest algorithm, proposed by Wager and Athey (2018) and Athey et al. (2019), to estimate the Conditional Average Treatment Effects (CATE):

$$\tau(x) = E[Y_i(1) - Y_i(0)|X_i = x].$$

This method is based on a standard regression tree: the algorithm starts with the whole (training) data set, takes a covariate, and splits the data into two leaves. The split is chosen such that it minimizes the goodness-of-fit criterion. The algorithm repeats this process until it reaches a terminal leaf. Within these terminal leaves, everyone shares values of certain covariates. Out-of-sample predictions are then made by determining which terminal leaf an observation belongs to, based on the covariates (Davis and Heller, 2017).

The CATE is obtained as the difference in the mean outcomes between a treatment and control observation within a terminal leaf (Davis and Heller, 2017). In other words, the CATE is the predicted treatment effect for out-of-sample observations that belong to a terminal leaf with specific values of a covariate.

When estimating the CATEs, we apply the so-called "honest" approach and grow so-called "honest trees" to obtain unbiased estimates and to ensure correct inference (Athey and Imbens, 2016; Wager and Athey, 2018): one part of the training data is used for building and growing the best fitting tree, i.e., it is used to estimate the model parameters and to determine the splits in the tree. The other part is used to estimate the treatment effects within each leaf of the tree using the estimated parameters. Hence, we fit two separate regressions.

Wager and Athey (2018) expand the idea of a causal tree to many trees: the Causal Forest, similar to the Random Forest algorithm proposed by Breiman (2001). Each tree is grown using greedy recursive partitioning on a random subset of the training data. A random split selection restricts the variables that are used at each split. Athey et al. (2019) preserve these elements in their Generalized Random Forest algorithm, but instead of averaging the estimates from each tree, they use a version of adaptive nearest neighbor estimator where "close" observations obtain more weight similar as in *k*-nearest neighbor estimations. More specifically, they use forest-based weights: in that case, a "close" observation is one which often ends up in the same leaf as the target value (Wager and Athey, 2018; Athey et al., 2020). In this case, the split criterion is to maximize the estimated treatment effect heterogeneity.

A4.2.2 Implementation

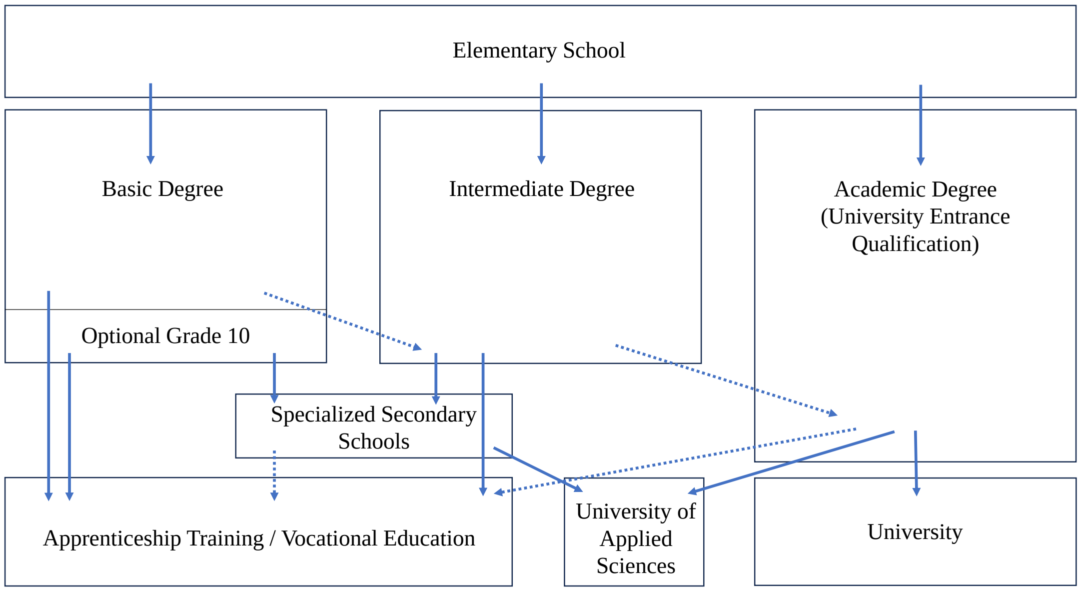
We include the following baseline characteristics in the estimation: age, female, born in Germany, West Germany, living in large city, risk, patience, parents with university education, income, current employment status (full-time, part-time, self-employed, unemployed, retired/ ill/etc.), middle school degree, university entrance degree, partner living in household, parental status, work in education sector, trust in government, education important for vote, general voting behavior.

We split the data set into 80 percent training and 20 percent test observations and evaluate the results on the test set.[1] We set the number of trees equal to 2,000 (according to the rule of thumb: number of observations). The number of variables that the algorithm examines at each split is set to five (square root of covariates).

[1] Honesty fraction = 0.5; minimum node size = 8.

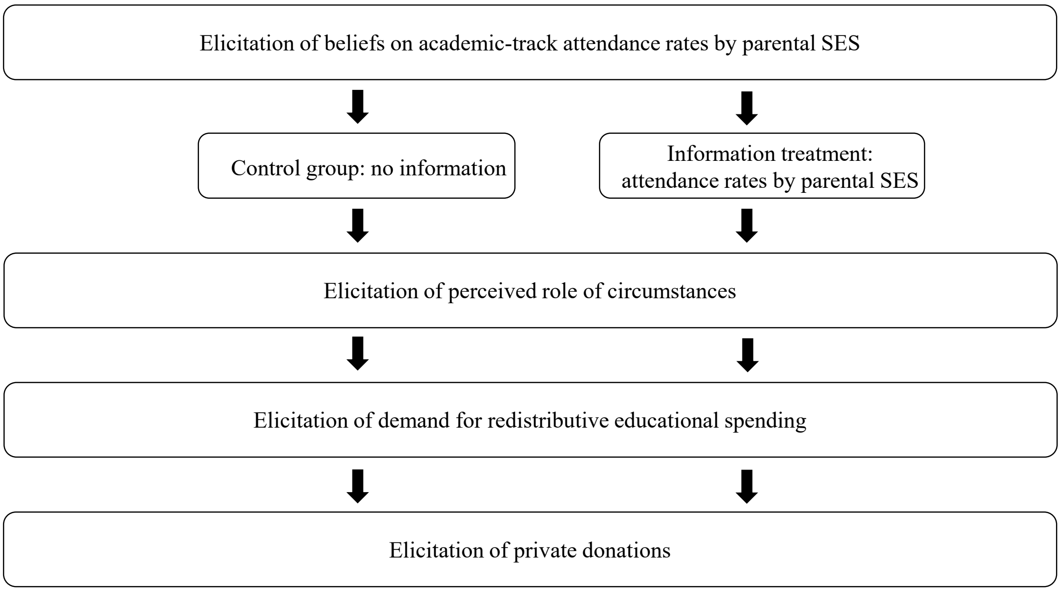
Appendix A4.3 Appendix Figures and Tables

Figure A4.1: The German Schooling System



Notes: The figure gives a schematic overview of the school system and degrees in Germany. After elementary school which takes four years (only in a few states six years), students are tracked into different school types where students can obtain the basic and intermediate degrees after grades 9 and 10, respectively. These degrees allow students to start apprenticeship training or other forms of vocational education. Students can also obtain the university entrance qualification after grade 13 (or 12). Switching tracks is, in principle, possible, enabling graduates from the basic and intermediate track to continue on to the next higher track, respectively, and/or obtaining their university entrance qualification via the specialized high track. Overview based on Grewenig (2021).

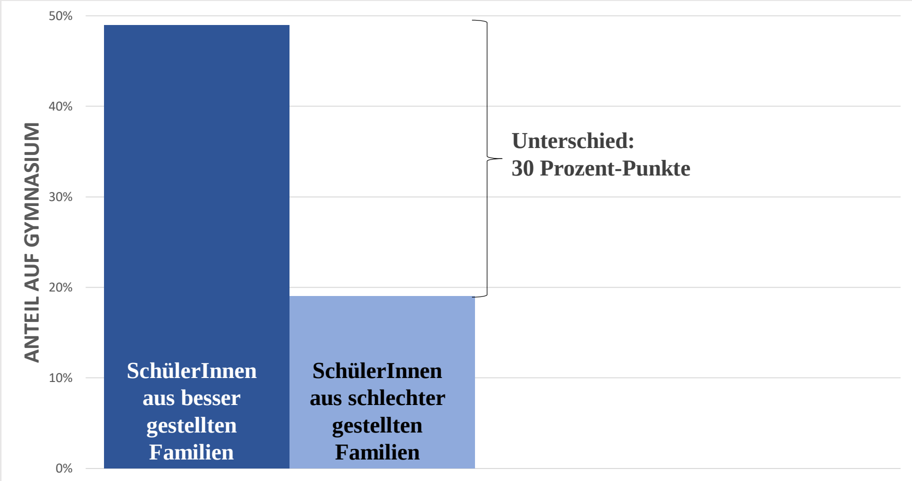
Figure A4.2: Experimental Design



Notes: The figure gives an overview of the experimental design. First, we elicit prior beliefs about the academic-track attendance rates of students from more and less advantaged backgrounds. Second, the treatment group is provided with the information while the control group is not. Third, we elicit three outcomes: the perceived role of circumstances for success (high educational degree and high income), the demand for redistributive education spending and the private donation decision.

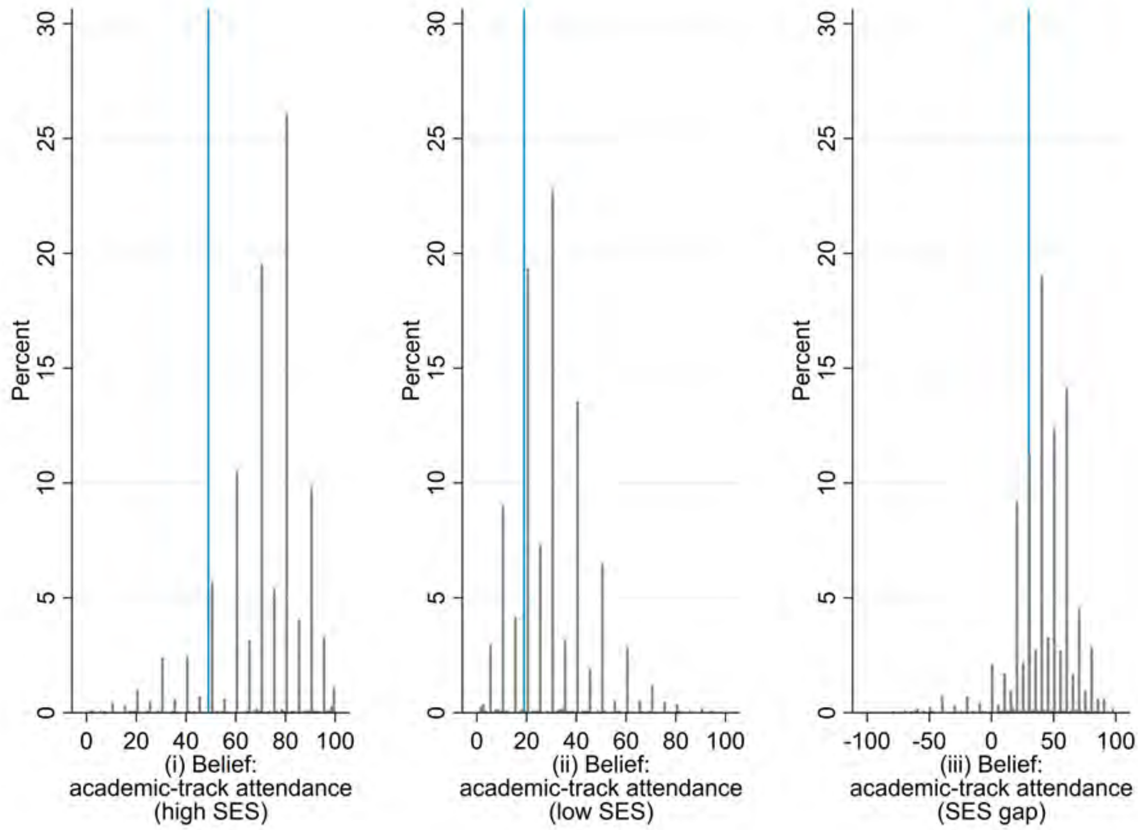
Figure A4.3: Illustration of the Information Treatment

49 Prozent der SchülerInnen aus der besser gestellten Hälfte aller Familien (in Bezug auf sozialen Hintergrund und familiäre Einkommensverhältnisse) besuchen ein Gymnasium. Unter SchülerInnen aus der schlechter gestellten Hälfte aller Familien sind es 19 Prozent. **Daraus ergibt sich ein Unterschied von 30 Prozentpunkten**.



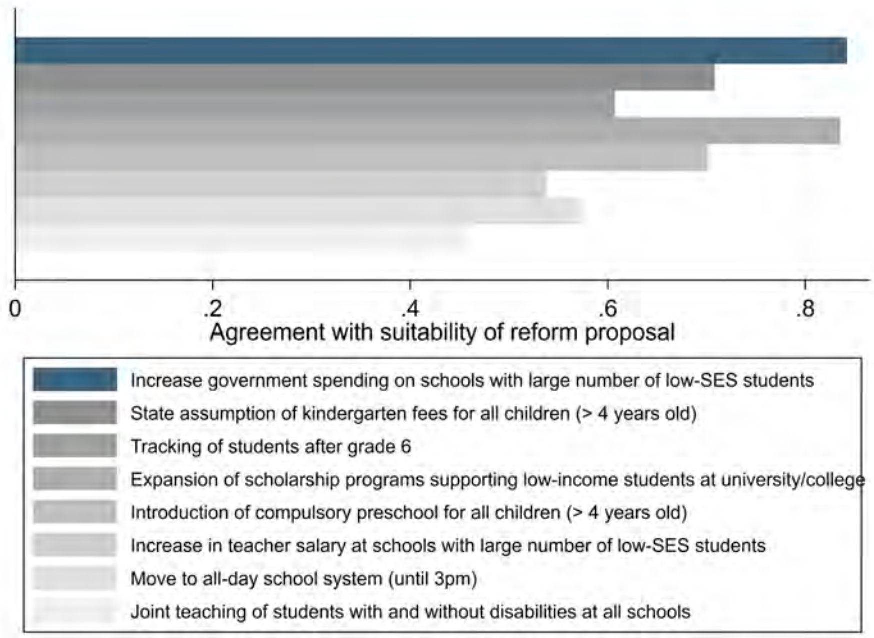
Notes: The figure shows the information that respondents in the treatment group were provided with (as they see it on the screen when they answer the survey). The information is provided in German. The English translation is the following: *49 percent of students from the better-off half of all families (in terms of social background and family income) attend a Gymnasium. Among students from the worse-off half of all families, the figure is 19 percent. This results in a difference of 30 percentage points.* Source: ifo Education Survey 2019.

Figure A4.4: Distribution of Prior Beliefs about Academic-Track Attendance by Parental Background

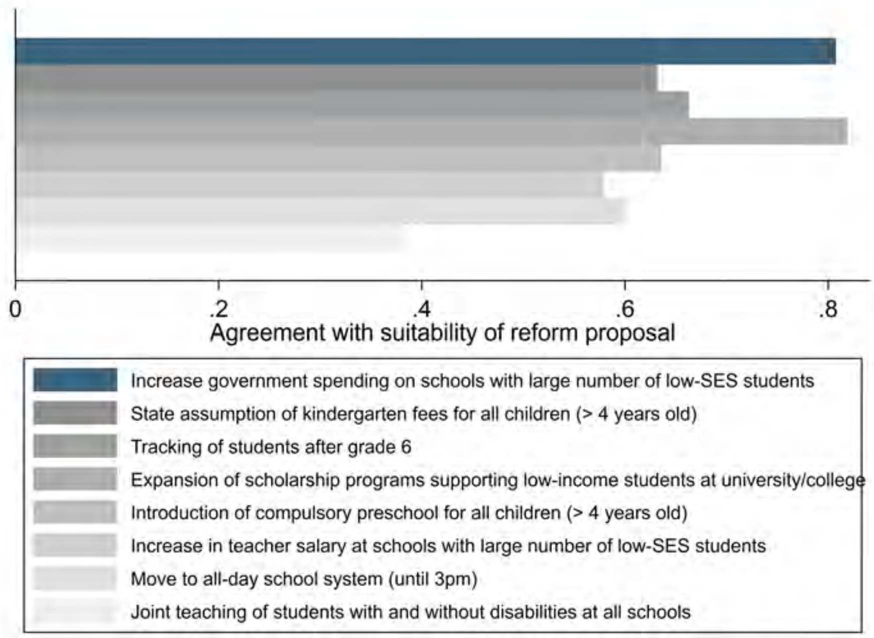


Notes: Histogram of the distribution of beliefs about academic-track attendance rates for students from different family backgrounds. Question wording: “Think of a comparison between children from the better- and worse-off half of all families (in terms of social background and family income). What do you think is the percentage of students from ... (i) the more advantaged half of all families who attend a *Gymnasium*?; (ii) the less advantaged half of all families who attend a *Gymnasium*?” The blue vertical lines indicate what would have been the correct answers. Data source: ifo Education Survey 2019.

Figure A4.5: Educational Reform Proposals in Germany



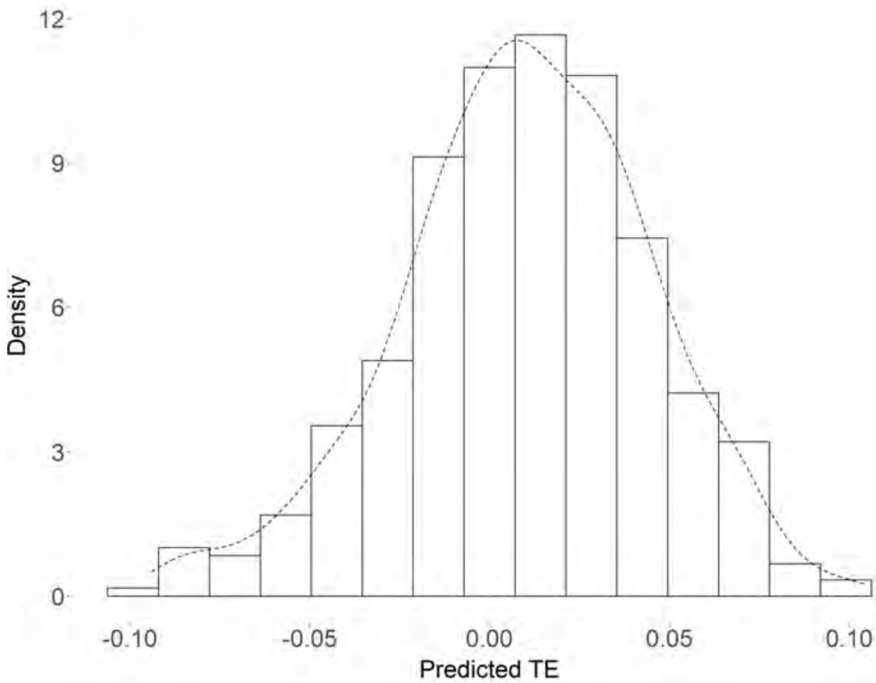
Panel A: Suitability of Reform Proposal to Increase Equality of Opportunity



Panel B: Suitability of Reform Proposal to Increase Average Student Performance

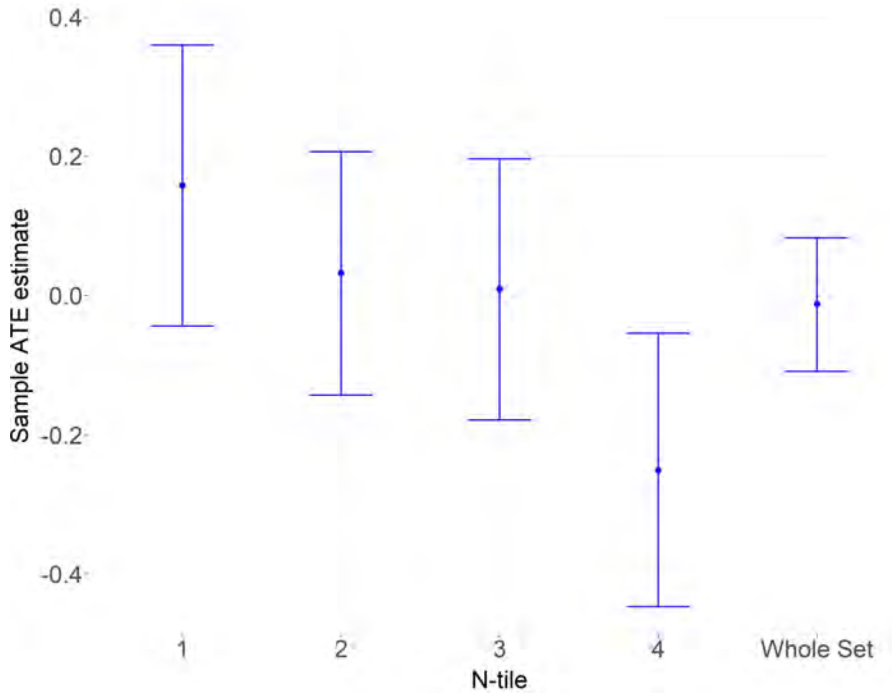
Notes: Question wording: “And how suitable do you think the reform proposals are for increasing equal opportunities in the German education system?” (Panel A); “And how suitable do you think the reform proposals are for raising the performance level in the German education system?” (Panel B). Data source: ifo Education Survey 2019.

Figure A4.6: Distribution of Conditional Average Treatment Effects (Demand for Redistributive Education Spending)



Notes: Distribution of the Conditional Average Treatment Effects for demand for redistributive education spending (on five-point scale). Data source: ifo Education Survey 2019.

Figure A4.7: Average Treatment Effects by N-tiles (Demand for Redistributive Education Spending)



Notes: Observations are split into four groups according to their predicted Conditional Average Treatment Effects. The figure shows the average treatment effect within these four groups and the whole sample for demand for redistributive education spending on five-point scale. Data source: ifo Education Survey 2019.

Table A4.1: Comparison of Analysis Sample to Microcensus Data

	Microcensus 2018 (1)	Analysis sample (2)
Age	50.941 (0.029)	53.067 (0.327)
Female	0.509 (0.001)	0.531 (0.011)
Living in West Germany (excl. Berlin)	0.803 (0.001)	0.796 (0.009)
Net household income above median	0.482 (0.001)	0.438 (0.011)
Educational attainment		
University entrance degree	0.341 (0.001)	0.413 (0.011)
Middle school degree	0.305 (0.001)	0.352 (0.010)
No degree / basic degree	0.354 (0.001)	0.234 (0.009)
Working full-time	0.438 (0.001)	0.323 (0.010)
Observations	445,867	2,094

Notes: Means; standard errors in parentheses. Column (1): all people aged 18 or older in the Microcensus 2018 (representative of the German population). Column (2): our analysis sample. Data sources: Microcensus 2018 and ifo Education Survey 2019.

Table A4.2: Respondent Characteristics across Treatment Arms

	Control Mean (1)	Treatment Mean (2)	Difference (3)	p-value (4)
Age	53.18	52.95	-0.24	0.72
Female	0.52	0.55	0.03	0.20
Born in Germany	0.95	0.96	0.01	0.54
City size $\geq$ 100,000	0.34	0.39	0.05	0.03
Partner in household	0.59	0.58	-0.01	0.64
Parent(s) with university degree	0.28	0.30	0.03	0.18
Highest educational attainment				
No degree/basic degree	0.23	0.24	0.01	0.71
Middle school degree	0.37	0.34	-0.03	0.15
Univ. entrance degree	0.40	0.42	0.02	0.29
Employment status				
Full-time	0.33	0.32	-0.01	0.78
Part-time	0.12	0.14	0.02	0.22
Self-employed	0.05	0.06	0.01	0.61
Unemployed	0.05	0.04	-0.01	0.32
Retired/ill/etc.	0.45	0.44	-0.01	0.72
Parent status	0.61	0.59	-0.02	0.44
Party preference				
CDU/CSU	0.17	0.19	0.01	0.46
SPD	0.18	0.15	-0.04	0.03
Grüne	0.14	0.16	0.01	0.47
Linke	0.10	0.10	0.00	0.86
FDP	0.06	0.05	-0.01	0.17
AfD	0.09	0.11	0.02	0.16
None	0.22	0.23	0.00	0.94
Other	0.02	0.02	0.01	0.22
Education important for vote	0.70	0.72	0.02	0.26
General voting	0.87	0.87	0.00	0.94
Patience	6.51	6.35	-0.16	0.10
Risk tolerance	4.60	4.74	0.14	0.22
Monthly household income (in EUR)	2556.21	2567.73	11.52	0.86
West Germany	0.79	0.80	0.01	0.52
Work in education sector	0.10	0.10	0.00	0.92
Trust in government	0.32	0.32	0.00	0.93

Notes: Group means. "Difference" displays the difference in means between the control group and the treatment group who received the information on the attendance rates. Data source: ifo Education Survey 2019.

Table A4.3: Participation in the Follow-Up Survey

	Respondent participated in follow-up survey	
	(1)	(2)
Treatment indicator		
Information on attendance rates	0.020	(0.017)
Covariates		
Age	0.005***	(0.001)
Female	0.013	(0.018)
Born in Germany	0.067	(0.049)
City size $\geq$ 100,000	-0.036*	(0.019)
Partner in household	-0.014	(0.021)
Parent(s) with university degree	0.009	(0.020)
Highest educational attainment		
No degree/basic degree (base category)		
Middle school degree	0.006	(0.021)
Univ. entrance degree	0.103***	(0.027)
Employment status		
Full-time	0.016	(0.047)
Part-time	-0.001	(0.050)
Self-employed	0.014	(0.057)
Unemployed (base category)		
Retired/ill/etc.	-0.015	(0.045)
Parent status	0.011	(0.021)
Party preference		
CDU/CSU (base category)		
SPD	-0.009	(0.029)
Grüne	0.002	(0.030)
Linke	-0.070*	(0.038)
FDP	0.024	(0.040)
AfD	-0.003	(0.037)
None	-0.011	(0.031)
Other	0.138***	(0.048)
Education important for vote	-0.014	(0.020)
General voting	0.004	(0.030)
Patience	-0.003	(0.004)
Risk tolerance	-0.004	(0.003)
Monthly household income (in EUR)	0.000	(0.000)
West Germany	-0.022	(0.022)
Working in education sector	-0.042	(0.030)
Trust in government	0.015	(0.020)
Observations	2,088	
R-squared	0.054	

Notes: Dependent variable: Dummy variable coded one if respondent participated in the follow-up survey. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.4: Respondent Characteristics across Treatment Arms in the Follow-up Sample

	Control Mean (1)	Treatment Mean (2)	Difference (3)	p-value (4)
Age	54.70	53.94	-0.76	0.27
Female	0.52	0.54	0.03	0.30
Born in Germany	0.96	0.96	0.00	0.85
City size $\geq$ 100,000	0.33	0.37	0.04	0.06
Partner in household	0.60	0.58	-0.02	0.50
Parent(s) with university degree	0.28	0.31	0.04	0.10
Highest educational attainment				
No degree/basic degree	0.21	0.22	0.01	0.47
Middle school degree	0.38	0.36	-0.02	0.29
Univ. entrance degree	0.41	0.42	0.01	0.67
Employment				
Full-time	0.33	0.31	-0.01	0.56
Part-time	0.12	0.13	0.02	0.33
Self-employed	0.05	0.06	0.01	0.38
Unemployed	0.05	0.04	-0.01	0.59
Retired/ill/etc.	0.46	0.45	-0.01	0.77
Parent status	0.63	0.61	-0.02	0.42
Party preference				
CDU/CSU	0.18	0.19	0.01	0.77
SPD	0.18	0.15	-0.03	0.10
Grüne	0.15	0.15	0.00	0.84
Linke	0.09	0.10	0.01	0.36
FDP	0.07	0.05	-0.01	0.27
AfD	0.10	0.11	0.01	0.54
None	0.21	0.22	0.01	0.71
Other	0.02	0.03	0.01	0.12
Education important for vote	0.69	0.72	0.03	0.14
General voting	0.88	0.87	-0.01	0.63
Patience	6.44	6.37	-0.08	0.47
Risk tolerance	4.51	4.74	0.24	0.07
Monthly household income (in EUR)	2613.71	2582.28	-31.43	0.68
West Germany	0.79	0.79	0.01	0.69
Work in education sector	0.09	0.09	0.00	0.99
Trust in government	0.33	0.32	-0.02	0.49

Notes: Group means. “Difference” displays the difference in means between the control group and the treatment group who received the information on the attendance rates. Sample: Follow-up survey participants. Data source: ifo Education Survey 2019.

Table A4.5: Effect of Information Treatment on Role of Circumstances and Effort: Robustness of Outcome Coding

	Circumstances decisive	Mainly circumstances	Rather circumstances	Rather effort	Mainly effort
	(1)	(2)	(3)	(4)	(5)
Panel A: High educational degree					
Information on attendance rates	0.123***	0.033***	0.090***	-0.023	-0.100***
	(0.018)	(0.008)	(0.017)	(0.022)	(0.020)
Covariates	Yes	Yes	Yes	Yes	Yes
Control mean	0.173	0.020	0.153	0.435	0.391
Observations	2,094	2,094	2,094	2,094	2,094
R-squared	0.060	0.019	0.048	0.014	0.053
Panel B: High income					
Information on attendance rates	0.036*	0.011	0.025	-0.036*	-0.000
	(0.021)	(0.011)	(0.020)	(0.021)	(0.018)
Covariates	Yes	Yes	Yes	Yes	Yes
Control mean	0.350	0.058	0.291	0.424	0.227
Observations	2,093	2,093	2,093	2,093	2,093
R-squared	0.040	0.018	0.026	0.021	0.029

Notes: OLS regressions. Dependent variables: (1) Dummy variable coded one if respondent thinks that mainly/rather external circumstances are decisive; (2) – (5) Dummy variable coded 1 = answer category given in respective table header, zero otherwise. Randomized experimental treatment “information on attendance rates”: respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.6: Effect of Information Treatment on Perceived Role of Circumstances, Private Donations, and Demand for Redistributive Education Spending (Two Treatment Groups)

	Perceived role of circumstances (high educ. degree)	Perceived role of circumstances (high income)	Average donations	Support for increased redistributive educ. spending
	(1)	(2)	(3)	(4)
Information on attendance rates (1^{st} *treatment group*)	0.258***	0.052	3.213**	-0.010
	(0.035)	(0.037)	(1.400)	(0.019)
Information on attendance rates (2^{nd} *treatment group*)	0.158***	0.050	2.906**	-0.018
	(0.034)	(0.038)	(1.420)	(0.019)
Covariates	Yes	Yes	Yes	Yes
Control mean	1.802	2.181	37.499	0.751
Observations	3,082	3,081	3,076	3,082
R-squared	0.055	0.036	0.047	0.032
Wald test for equality of treatment effects [1^{st} *treatment group* = 2^{nd} *treatment group*]				
p-values	0.006	0.957	0.826	0.690

Notes: OLS regressions. Dependent variables: (1) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances; (2) Perception that circumstances, not effort, are decisive for high income on a four-point scale, with higher values indicating more importance of circumstances; (3) Amount of donations stated by respondents (in tokens); (4) Dummy variable coded one if respondent is mainly/rather in favor of redistributive education spending. Randomized experimental treatment "information on attendance rates (1^{st} *treatment group*)": respondents informed about differences in academic-track attendance rates by parental background. Randomized experimental treatment "information on attendance rates (2^{nd} *treatment group*)": respondents informed about differences in academic-track attendance rates by parental background conditional on math and reading achievement. Control mean: mean of the outcome variable in the control group. See Table 4.1 for included covariates. Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.7: Effect of Information Treatment on Perceived Role of Circumstances, Private Donations, and Demand for Redistributive Education Spending by Educational Attainment

	Perceived role of circumstances (high educational degree)		Average donations		Support for increased redistributive educ. spending	
	No academic-track school degree	Academic-track school degree	No academic-track school degree	Academic-track school degree	No academic-track school degree	Academic-track school degree
	(1)	(2)	(3)	(4)	(5)	(6)
Information on attendance rates	0.231***	0.279***	4.981***	-0.801	-0.012	-0.019
	(0.047)	(0.054)	(1.856)	(2.240)	(0.025)	(0.030)
Covariates	No	No	No	No	No	No
Difference b/w groups	0.048		-5.782**		0.007	
Control mean	1.752	1.876	34.739	41.622	0.749	0.755
Observations	1,230	864	1,229	864	1,230	864
R-squared	0.019	0.030	0.006	0.000	0.000	0.000

Notes: OLS regressions. Dependent variables: (1) – (2) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances; (3) – (4) Amount of donations stated by respondents (in tokens); (5) – (6) Dummy variable coded one if respondent is mainly/rather in favor of redistributive education spending. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable for the control group. See Table 4.1 for included covariates. Difference b/w groups: difference between the two groups (no academic-track school degree vs. academic-track school degree). Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.8: Effect of Information Treatment on Perceived Role of Circumstances, Private Donations, and Demand for Redistributive Education Spending by Political Ideology

	Perceived role of circumstances (high educational degree)			Average donations			Support for increased redistributive educ. spending		
	Left-leaning	Right-leaning	No attachment	Left-leaning	Right-leaning	No attachment	Left-leaning	Right-leaning	No attachment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Information on attendance rates	0.333***	0.180***	0.258***	3.424	2.847	1.453	-0.010	-0.014	0.013
	(0.057)	(0.057)	(0.074)	(2.218)	(2.444)	(2.983)	(0.024)	(0.035)	(0.044)
Covariates	No	No	No	No	No	No	No	No	No
Difference left vs. right		0.153*			0.577			-0.004	
Difference left vs. no attachm.		0.075			1.971			0.023	
Difference right vs. no attachm.		-0.078			1.394			0.027	
Control mean	1.880	1.718	1.750	41.390	35.161	33.428	0.862	0.692	0.627
Observations	868	710	472	867	710	472	868	710	472
R-squared	0.038	0.014	0.025	0.003	0.002	0.001	0.000	0.000	0.000

Notes: OLS regressions. Dependent variables: (1) – (3) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances; (4) – (6) Amount of donations stated by respondents (in tokens); (7) – (9) Dummy variable coded one if respondent is mainly/rather in favor of redistributive education spending. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable for the control group. See Table 4.1 for included covariates. Difference b/w groups: difference between the three groups (left-leaning vs. right-leaning vs. no attachment). Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.9: Effect of Information Treatment on Perceived Role of Circumstances, Private Donations, and Demand for Redistributive Education Spending by Trust in Government

	Perceived role of circumstances (high educational degree)		Average donations		Support for increased redistributive educ. spending	
	Low trust in government	High trust in government	Low trust in government	High trust in government	Low trust in government	High trust in government
	(1)	(2)	(3)	(4)	(5)	(6)
Information on attendance rates	0.285***	0.190***	1.965	4.164*	-0.021	-0.003
	(0.044)	(0.060)	(1.739)	(2.470)	(0.024)	(0.031)
Covariates	No	No	No	No	No	No
Difference b/w groups	-0.095		2.199		0.018	
Control mean	1.789	1.830	35.150	42.491	0.730	0.798
Observations	1,422	672	1,421	672	1,422	672
R-squared	0.029	0.015	0.001	0.004	0.001	0.000

Notes: OLS regressions. Dependent variables: (1) – (2) Perception that circumstances, not effort, are decisive for high educational attainment on a four-point scale, with higher values indicating more importance of circumstances; (3) – (4) Amount of donations stated by respondents (in tokens); (5) – (6) Dummy variable coded one if respondent is mainly/rather in favor of redistributive education spending. Randomized experimental treatment "information on attendance rates": respondents informed about differences in academic-track attendance rates by parental background. Control mean: mean of the outcome variable for the control group. See Table 4.1 for included covariates. Difference b/w groups: difference between the two groups (low vs. high trust in government). Data source: ifo Education Survey 2019. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.10: Covariates by N-tiles (Demand for Redistributive Education Spending)

Covariates	N-tile1	N-tile2	N-tile3	N-tile4	p-value
Middle school degree	0.33	0.43	0.39	0.26	0.02
Univ. entrance degree	0.50	0.34	0.36	0.50	0.97
Age	48.81	56.24	54.83	52.64	0.00
Monthly household income (in EUR)	2.38	2.47	2.50	2.94	0.00
Female	0.58	0.60	0.53	0.41	0.00
Born in Germany	0.93	0.96	0.95	0.99	0.00
Partner in household	0.58	0.58	0.54	0.67	0.01
West Germany	0.91	0.81	0.77	0.71	0.00
City size $\geq$ 100,000	0.30	0.33	0.38	0.39	0.00
Parent status	0.54	0.64	0.59	0.64	0.01
Parent(s) with university degree	0.35	0.26	0.30	0.28	0.02
Retired/ill/etc.	0.16	0.23	0.33	0.57	0.00
Full-time employed	0.17	0.16	0.10	0.05	0.00
Part-time employed	0.02	0.04	0.06	0.09	0.00
Self-employed	0.03	0.03	0.06	0.07	0.01
Unemployed	0.62	0.54	0.45	0.23	0.00
Work in education sector	0.08	0.09	0.10	0.11	0.12
Risk tolerance	4.91	4.82	4.70	4.30	0.00
Patience	7.36	6.90	6.34	5.14	0.00
Left-leaning party	0.16	0.27	0.33	0.31	0.00
No party preference	0.11	0.19	0.29	0.32	0.00
Right-leaning party	0.53	0.38	0.23	0.23	0.00
Trust in government	0.39	0.34	0.28	0.28	0.00
Education important for vote	0.70	0.72	0.71	0.69	0.80
General voting	0.89	0.88	0.87	0.85	0.08
No degree/basic degree	0.00	0.00	0.00	0.00	1.00

Notes: Variables included in the Causal Forest estimations (demand for redistributive education spending). Mean value of variables for four groups split according to the predicted Conditional Average Treatment Effect. p-value for difference between first and fourth group. Data source: ifo Education Survey 2019.

Table A4.11: Variable Importance (Demand for Redistributive Education Spending)

Covariates	Variable importance
Continuous variables	
Age	0.13
Patience	0.10
Monthly household income (in EUR)	0.10
Risk tolerance	0.08
Binary variables	
West Germany	0.04
Unemployed	0.04
Education important for vote	0.04
Female	0.03
Retired/ill/etc.	0.03
Full-time employed	0.03
Middle school degree	0.03
Parent(s) with university degree	0.03
City size $\geq$ 100,000	0.03
Parent status	0.03
Right-leaning party	0.03
Partner in household	0.03
No party preference	0.03
Left-leaning party	0.03
Univ. entrance degree	0.03
General voting	0.02
Part-time employed	0.02
Trust in government	0.02
Work in education sector	0.02
Born in Germany	0.01
Self-employed	0.01
No degree/basic degree	0.00

Notes: Variable importance measure from the Causal Forest (demand for redistributive education spending). The maximum tree depth when calculating the variable importance is four. Data source: ifo Education Survey 2019.

5 Automatability of Occupations, Workers' Labor-Market Expectations, and Willingness to Train*

5.1 Introduction

Technological progress is a key driver of economic growth (Aghion and Howitt, 1992; Jones, 2002; Acemoglu, 2009). Changes in technology often bring about drastic changes in the demand for different input factors. One notable example is the ongoing digital transformation: technologies like artificial intelligence have already transformed the skill demand in many professions, rendering existing skills and sometimes entire professions obsolete (OECD, 1998; Dachs, 2018; OECD, 2021). At the same time, there is an increasing demand for occupations and skills that complement the new technologies.[1] Because of these new and changing technological opportunities and the rising automatability of occupations, the knowledge that workers have acquired becomes outdated at an ever-faster rate. Therefore, continued learning throughout one's working life through ongoing further training becomes crucial to keep pace with structural change on the labor market and to sustainably benefit from structural change in the economy in the long term (e.g., Bessen, 2019; Innocenti and Golin, 2022).

In this paper, we focus on the question of whether individuals are aware of their occupations' automatability and how it influences their decision to participate in further training and their labor-market expectations. More specifically, since a significant portion of the German working-age population barely engages in further training activities (OECD, 2021), we are interested in the role of automatability as a driver for this inequality in participating in further training across occupations. Individuals in occupations with a high automatability are strongly underrepresented in continued education and further training initiatives (Heß et al., 2019; OECD, 2021) whereas their employment prospects would benefit most from further training.[2] This inequality in participating in further training between individuals in high and low automatability occupations is not yet fully understood and therefore presents an important research gap that we address in this paper. We hypothesize that one contributing factor to the low engagement in further training among workers in occupations with a high

* This chapter is co-authored with Philipp Lergetporer and Katharina Werner.

1 Bundesministerium für Arbeit und Soziales (2021) estimates that 5.3 million jobs will be lost in Germany by 2040, while 3.6 million new jobs will be created.

2 Several studies show that workers in high-automatability occupations have worse employment outcomes and lower wage growth, which highlights the importance of the results of our study for this group of workers (Acemoglu and Restrepo, 2020b; Dauth et al., 2021; Georgieff and Milanez, 2021; Schmidpeter and Winter-Ebmer, 2021; Montobbio et al., 2022).

automatability is that they hold biased beliefs about their occupations' automatability.[3] Existing research has demonstrated that people often have misperceptions about labor-market relevant information such as the probability of finding a job during unemployment, wages, and outside options (see e.g., Jäger et al., 2021; Mueller et al., 2021). If workers underestimate their occupations' automatability, they may underinvest in ensuring their skills are up-to-date for changing labor-market requirements. Given the substantial number of workers in Germany with a high automatability, which has grown in the past (Dengler and Matthes, 2018a), better understanding their further training decisions becomes crucial.

While previous literature has focused on the expected effects of automation on the German labor market, very little is known about workers' own beliefs regarding their occupation's automatability. We address this gap by (i) documenting beliefs about automatability in an online sample of the German population and (ii) studying experimentally how providing information about their occupations' automatability affects workers' beliefs, labor-market expectations, and preferences regarding further training. By examining the effects of information provision, we can shed light on the potential impact of correcting biased beliefs on individuals' decision-making processes and their engagement in further training.

For this experiment, we conducted a large online survey (N = 3,012) to represent the German adult population. In the survey, we first elicited respondents' prior beliefs about the automatability of their occupation. We then provided a randomly selected treatment group with personalized, occupation-specific information about the automatability of each respondent's occupation based on estimates by the Research Institute of the Federal Employment Agency (IAB). The remaining participants served as an uninformed control group. For the information treatment, we draw on information on the automatability of occupations from the IAB-Job-Futuromat[4] and provide respondents in the treatment group with the respective number on the share of automatable core tasks in their occupation from the IAB-Job-Futuromat, i.e., their occupation's automatability. An occupation's automatability is defined as the share of automatable core tasks among all core tasks within an occupation (Dengler and Matthes, 2018a). Automatable in this context means that job tasks could theoretically be carried out by a computer or fully automatically by a computer or computer-controlled machine. Finally, we measure (i) respondents' labor-market expectations about their professional future, (ii) respondents' likelihood of participating in further training and retraining as well as (iii) the extent to which respondents are willing to forgo a fraction of their wage during the period of further training. By comparing responses between the uninformed control and the informed treatment group, we evaluate how factual information about the automatability of workers' occupations affects these outcomes.

---

[3] Complementary reasons for non-participation in further training include, e.g., stigmatization (since further training possibilities are sometimes offered by the Federal Employment Agency, which most people in Germany associate with unemployment), failure to recognize potential benefits, high costs, or lack of time resources (e.g., van den Berg et al., 2019; Müller and Wenzelmann, 2020; Osiander and Stephan, 2020). We report further details on barriers to participation in section 5.4.3.

[4] The IAB-Job-Futuromat can be accessed online at `https://job-futuromat.iab.de/`.

We find that, on average, respondents underestimate the automatability of their own occupation. Crucially, the misperception is particularly large for those in occupations with a high automatability. Descriptively, we show that in the control group, respondents who believe that very few of the tasks in their occupation are automatable tend to state the lowest likelihood of participating in further training. Furthermore, there is a positive correlation of respondents' beliefs about their occupations' automatability and their plans to participate in retraining. The information experiment allows us to investigate whether exogenous shifts in beliefs about automatability do in fact have a causal effect on the outcome variables.

Experimental results show that information provision about the automatability of one's occupation affects labor-market expectations. Respondents become 9.7 percent of a standard deviation more concerned about their future work and are more likely to expect changes in their work environment (13.0 percent of a standard deviation). Furthermore, providing information increases the stated likelihood of participating in further training and retraining by 5.8 percent and 12.5 percent, respectively, relative to the mean. It also results in an increase in the fraction of wages that respondents are willing to forgo to participate in further training.

In line with our findings on misperceptions, the treatment effects for respondents in occupations with high automatability are larger than those with a low automatability. Control group respondents in occupations with a high automatability express a 37.6 percent likelihood that they will participate in further training. Treated respondents in occupations with a high automatability state a 4.6 percentage points greater likelihood of participation. In contrast, the treatment effect for those in occupations with a low automatability is minimal, amounting to -0.1 percentage points. Similarly, treated respondents in a high-automatability occupation state a 5.2 percentage points greater likelihood of participating in retraining while the treatment effect for those in low automation occupations is 1.5 percentage points. Thus, information provision reduces the gap in the willingness to participate in further training by 95.5 percent and completely closes it for respondents' willingness to participate in retraining. The same pattern can be observed regarding the willingness to forgo wage. These results highlight that a lack of information about the occupation's automatability could be an important factor in explaining the documented low participation in further training among those who work in occupations most likely to be affected by technological change and automation.

Further results show that reluctance to participate in further training is not due to ignorance of the benefits. A large majority (76.7 percent) of respondents agree that further training is useful for keeping pace with structural change. Similarly, 66.2 percent agree that the future need for further training will increase for all employees and 62.5 percent agree that everyone affected by structural change should participate in further training. Respondents state that the main reasons for not participating in further training are financial constraints (45.4 percent), lack of employer support (45.0 percent), and time constraints (35.2 percent). Answers on these outcomes are unaffected by treatment group status, i.e., do not change with information on respondents' occupations' automatability. Treated respondents also request additional information about further training programs and finance options at the same rate as respondents in the control group.

The remainder of this paper is structured as follows. Section 5.2 provides an overview of the literature. Section 5.3 presents the experimental setting and the data. Section 5.4 presents descriptive and experimental results. Section 5.5 concludes.

5.2 Related Literature

Our study builds upon several strands of the existing literature. First, we relate to the literature on the effects of technology and automation on labor-market outcomes. Most previous studies in this field have predominantly used non-survey data to examine the impact of automation on employment and wages. Acemoglu and Restrepo (2018, 2020a) identify two main theoretical impacts of new technologies on employment and wages: the displacement effect, implying that robots take over tasks previously performed by humans, and the productivity effect, which implies gains in productivity, thus increasing the demand for labor in non-automated occupations (Acemoglu and Restrepo, 2019) and leading to new occupations and tasks.[5] Several studies have empirically examined the effect of (industrial) robots on employment outcomes (Autor and Salomons, 2018; Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020a; Dauth et al., 2021). While most studies find no effect on total employment and a positive impact on productivity and wages (Graetz and Michaels, 2018; Dauth et al., 2021), Acemoglu and Restrepo (2020a) find negative effects of robots on employment and wages across commuting zones in the U.S. The overall null effect on total employment is often masked by displacement and re-allocation effects (Dauth et al., 2021), which is why automation and digitization do not necessarily lead to negative consequences on (aggregate) employment (Bessen, 2019; Arntz et al., 2020). In light of these findings, some researchers argue that automation is more likely to lead to an increase in worker transitions than mass unemployment (Bessen et al., 2020). However, either in the case of workplace transition or unemployment, workers must retrain, frequently acquiring new skills and changing occupations and industries. Further training programs are therefore vital for ensuring continued matching of labor demand and supply in rapidly changing labor markets. We contribute to this literature by using survey data and focusing specifically on the link between workers' beliefs about automatability and their labor-market expectations, as well as their willingness to (re)train.

Secondly, our research is in line with studies that examine the type of workers vulnerable to replacement and changes to their occupation-specific required skills. For example, Blanas et al. (2019) argue that some workers respond to automation by transitioning to low-paid

[5] Recent studies on the emergence of "new work" and the task-based approach discuss how automation and computerization reallocate many human tasks to machine tasks, expanding the set of tasks performed by capital (Acemoglu and Restrepo, 2022), and how they complement educated workers (Autor, 2022). Consequently, the introduction of new job tasks or job categories requires specialized human expertise, requiring more education and training of the workforce (Autor, 2022; Autor et al., 2022).

occupations where tasks are difficult to replace with machines, while other workers acquire new skills that complement machines and allow them to work in high-paid occupations. In addition, Cortes (2016) examines the effects of routine-biased technological change[6] on workers' occupational transitions and their wage changes, finding that especially low-ability routine workers tend to switch to non-routine manual tasks, while high-ability routine workers switch to non-routine cognitive occupations. We contribute to this literature by investigating one potential reason for the lower training participation rates of at-risk individuals and by exploring the occupations they consider in case of retraining.

Other studies have examined barriers to training participation, such as misjudgment of potential benefits, high costs, or lack of time resources (e.g., van den Berg et al., 2019; Müller and Wenzelmann, 2020; Osiander and Stephan, 2020). We complement this by focusing on differences in participation rates across subgroups of workers and examining the influence of beliefs regarding their automatability on training participation.

Moreover, our study also relates to the literature on the relationship between automation and further training. Heß et al. (2019) find that participation in further training strongly differs depending on the proportion of routine tasks in workers' occupations: employees with the highest share of routine tasks have the lowest participation rates (27 percent), while 41 percent of employees with low shares of routine tasks participated in further training courses. Innocenti and Golin (2022) show that 30 percent of respondents in their sample from several countries are worried about being replaced by machines or algorithms and that workers' fear of automation is positively associated with their intentions to invest in training activities. Nedelkoska and Quintini (2018) use data from the Programme for the International Assessment of Adult Competencies (PIAAC) assessment and find that workers in occupations at risk of automation have lower on-the-job and outside training participation rates. They show that the likelihood of having not participated in on-the-job training during a twelve-month period is three times higher for those working in occupations that are fully automatable compared to those working in non-automatable jobs. While these studies are primarily descriptive, we add to the literature by examining the causal link between perceptions about the automatability and participation in further training.

In terms of methodology, our paper contributes to the literature that uses survey experiments to examine the effect of information provision on public preferences (e.g., Cruces et al., 2013; Kuziemko et al., 2015; Bursztyn, 2016). The experimental literature on information provision about automation and the future of work is closest to our study. Jeffrey (2021) provides information about the perceived vulnerability to an automation shock, namely that more jobs are lost than gained and that a family member loses a job, focusing on redistributive

[6] Since the 1980s, technological changes have led to the occurrence of more machines and computers that mainly perform routine tasks. Therefore, mainly routine workers are substituted. This hypothesis is termed routine-biased technological change (Cortes, 2016). Earlier literature hypothesized that technological change is skill-biased, which favored high-skilled workers without distinguishing between tasks and skills (e.g., Katz and Murphy, 1992; Autor et al., 1998).

preferences as an outcome. She finds no effect on redistributive preferences on average. On the other hand, support for redistributive policies increases when rhetoric is added that leads respondents to perceive automation-induced disparity as unfair. Arntz et al. (2022) provide two pieces of scientific information about labor-market effects of automation (i.e., no aggregate employment losses, employment shift from unskilled to skilled workers). They find that information about zero net employment effects reduces concerns about automation, but they find no average effect on stated labor-market behavior and donations to NGOs. These findings are heterogeneous across respondents with different prior beliefs. In comparison to these studies, our contribution lies in providing respondents with personalized, occupation-specific information about the automatability of their occupation. A study by Golin and Rauh (2022) in the U.S. where respondents are informed about the expected probability of job loss for their current occupation comes closest to our paper. They find a causal effect of information provision on preferences for redistribution and respondents' stated likelihood of joining a worker union, but no effect on intentions to participate in retraining or switching occupations. In contrast to their study, we focus on estimates of the automatability of core tasks within an occupation, rather than unemployment risk. Given recent advances in the literature, understanding expectations of changes in the labor market beyond unemployment is especially relevant to the German context as well as to many developed countries that experience worker shortages. Our main focus on labor-market outcomes, specifically on participation in further training and retraining, is particularly relevant in a context where matching of worker (skills) to tasks is currently emerging as a large societal challenge.

5.3 Experimental Setting

In this section, we introduce the institutional background of further training in Germany (section 5.3.1), the data from the ifo Education Survey (section 5.3.2), the survey experiment (section 5.3.3), and describe the sample (section 5.3.4).

5.3.1 Institutional Background

In Germany, the majority of further training occurs within the company context: 72 percent of all further training activities take place within the company (Bundesministerium für Bildung und Forschung, 2019). The average duration of in-company further training is 29 hours (per training), which is shorter than the average individual job-related training (outside the company), amounting to approximately 153 hours (per training) (Bundesministerium für Bildung und Forschung, 2019). According to the ifo Education Survey 2022, 63 percent of respondents had participated in further training in the past, indicating that more than one third (37 percent) had not yet engaged in any form of additional training (Werner et al., 2022).

Overall, there are approximately 18,000 public and private further training providers (Bundesinstitut für Berufsbildung, 2020) and the sector operates under various levels of regulation and numerous legal bases, including collective bargaining agreements, company agreements, laws, and state-level regulations. The responsibility and financial burden for further training are shared among companies, workers, and the public sector. The latest law on further training in Germany introduced in 2020, the "work-of-tomorrow law" (*Arbeit-von-morgen-Gesetz*), regulates that further training of more than 120 hours can be subsidized by the Federal Employment Agency. Depending on the size of the company, up to 100 percent of the course costs can be financed through the Federal Employment Agency. This law is specifically aimed at people with occupations that can be replaced by technology or are otherwise affected by structural change. For more information about the institutional background, see Appendix A5.1.

5.3.2 Data

We use data from the ifo Education Survey 2022, a large opinion survey on education policy in Germany (Freundl et al., 2023). The survey company Talk Online conducted sampling and polling in May/June 2022. Overall, the survey encompassed questions related to education policy, focusing on topics related to structural change[7] and lifelong learning. In addition to this, respondents were asked about an extensive set of socio-demographic background characteristics at the end of the survey. The median completion time was 16 minutes. Moreover, the item non-response rate is low, with a maximum of 2.9 percent for the questions we use in this study. We restrict our sample to respondents who state that they are currently employed. Since the sample was drawn up to reflect the German population using quotas for gender, age, state, education level, and employment status, our data cover a broad sample of the German working population from different occupational fields with different requirement levels (see section 5.3.4). The overall sample size is 3,012 respondents.

Respondents were sampled and surveyed online, which means they answered the survey autonomously on their digital devices. On the online platform, respondents can take surveys in exchange for rewards. In our survey, all respondents are incentivized for survey completion. This money can either be paid out in cash or for vouchers at different retailers.[8]

5.3.3 The Experiment

We conducted a survey experiment that provides respondents in the treatment group with personalized information about the automatability of their occupation. The experimental setup is as follows (see also Appendix Table A5.1): first, we elicit respondents' current occupation and their beliefs about the automatability of their occupation. Next, we provide respondents in the treatment group with information about their occupation's automatability before eliciting the main outcomes, i.e., labor-market expectations, likelihood of participating

[7] We define structural change to respondents in the following way: *By structural change, we mean the constant transformation of economic sectors accelerated by digital technologies, among other things.*

[8] Respondents obtained 1.50€ for their participation in our survey.

in further training and retraining, and the wage fraction they are willing to forgo during further training. Respondents in the control group answer the same questions without receiving the information. Finally, all respondents are asked about potential barriers to participating in further training.

Information Treatment

We hypothesize that underestimation of the automatability of tasks contributes to low participation rates in further training for some subpopulations of workers. Workers might have biased beliefs about the displacing effects of technologies and, therefore, about the impact and extent of automation on the labor market and their own occupation. We thus provide treatment group respondents with information about their occupation's automatability in our survey.

At the beginning of the survey, we ask all respondents which occupation they currently work in.[9] According to the respondents' occupation, we then provide personalized information about the automatability of this occupation to the randomly selected treatment group. We use information on the automatability calculated by the IAB, which can be accessed through the IAB-Job-Futuromat, a public website.[10] The automatability ranges from zero to 100 percent. The information that we provide reads as follows: *"According to a study, [X] percent of core tasks in occupation [answer from earlier question about current occupation] are as of today automatable."* The values in brackets are adjusted for each respondent according to his or her current occupation.[11] Along with the verbal statement, we also provide respondents with a graphical visualization of the information (see Appendix Figure A5.1). Previous research

[9] Respondents can choose their occupation from a list of more than 4,000 occupations included in BERUFENET. BERUFENET is an expert database for training and job descriptions from the German Federal Employment Agency. It is similar to the U.S. Occupational Information Network (O*NET).

[10] In order to obtain this measure of automatability, the IAB relies on a list of core tasks that are typically done by workers in each occupation. Subsequently, experts code for each of these tasks whether they could be performed fully automated with currently available technology (for details, see Dengler and Matthes (2018a) and `https://job-futuromat.iab.de/faq.html`).

[11] For the information treatment, we use information on the *general* occupation-specific automatability. This assumes that a person spends the same amount of time on each task in an occupation (Dengler and Matthes, 2018b). Dengler and Matthes (2018b) argue that on the one hand, this could lead to an overestimation of the automatability if less time is spent on tasks with a high share of automatable core tasks. On the other hand, it could lead to an underestimation if a person spends more time on these tasks. In our study, we present the general occupation-specific automatability even though respondents might perform automatable tasks to a higher or lower extent and the automatability might vary within occupations. The information that we provide therefore states the automatability of a specific occupation, even repeating the name of the occupation, and does not mention that it is the automatability of respondents' own job.

shows that this measure of automatability has important predictive effects on employment growth (Dengler and Matthes, 2018b).[12]

Eliciting Beliefs about the Automatability

We elicit respondents' perceptions of the automatability of (i) their occupation in general and (ii) their own job. Prior to the information treatment, we first ask respondents about the share of core tasks they think are automatable in their current occupation in order to assess their beliefs about their occupation's automatability at baseline.[13] We also offer respondents an example on how the automatable share of core tasks is calculated through a further click on the screen. In addition, we elicit respondents' confidence in their beliefs on a seven-point Likert scale.

Second, after providing information to the treatment group, we elicit all respondents' beliefs about the share of automatable core tasks that they perform themselves in their jobs.[14] We explicitly phrased this question slightly differently to the previous question since this might have irritated respondents and led to problems in answering behavior. The distinction between occupation-wide beliefs and beliefs regarding their own job allows us to investigate to what extent respondents in the treatment group update beliefs on their own share of automatable tasks when they receive information relating to the occupation average. This is particularly important as respondents are likely to have private information on the type of tasks that they perform in their current job, which will introduce unobserved heterogeneity in respondents' automatability within their occupation.

Eliciting Labor-Market and Further Training Outcomes

Our main outcomes are (i) respondents' labor-market expectations, (ii) respondents' stated likelihood that they will participate in further training and retraining, and (iii) the fraction of wages that respondents are willing to forgo during further training.

First, we elicit respondents' labor-market expectations by asking to what extent they agree with a number of statements regarding their professional future. The nine statements can be grouped into two domains: the first is to examine whether respondents are concerned about their professional future and about being replaced by computers or machines (*labor-market concerns*). The second domain is whether respondents expect changes in their work environment in the future, e.g., whether they will have different and/or more demanding tasks,

[12] The approach by Dengler and Matthes (2018a) is similar to the job-level (task-based) approach by Arntz et al. (2017), who estimate that about twelve percent of workers in Germany have an automation risk greater than 70 percent (Arntz et al., 2016). For the U.S., Arntz et al. (2016) calculate that approximately nine percent of workers work in jobs with an automation risk above 70 percent, which is lower than the occupation-level approach by Frey and Osborne (2017), who estimate that about 47 percent of jobs are at risk of automation.

[13] The belief elicitation question is worded as follows: *What do you think is the percentage of core activities that people perform in occupation [answer from earlier question about current occupation] that can be automated?*

[14] This question is worded as follows: *What percentage of the core activities that you specifically perform in your job do you think can be automated?*

and whether they will work fewer hours (*work-environment change*).[15] We combine items by domain into two indices which are standardized to mean zero and standard deviation one. Higher values of the index *labor-market concerns* reflect greater concerns about the professional future and higher values on the index *work-environment change* higher expectations that the work environment will change. Combining the nine items into two indices can alleviate concerns of multiple hypothesis testing and improve statistical power (Anderson, 2008; Heller et al., 2017).

Second, we elicit respondents' likelihood of participation in further training and retraining to another occupation in the next two years on a scale from zero to 100 percent. Further training courses are very heterogeneous in intensity. For the purpose of our study, we follow funding eligibility criteria from the Federal Employment Agency and focus on courses consisting of at least 120 hours of training (see section 5.3.1 for details on the institutional background). In the case of retraining, we refer to a further training program in which respondents acquire skills for a new occupation, rather than acquiring skills that can be applied in their current job. In addition to the likelihood of participating in a retraining program, we also ask respondents which occupation they would retrain in if they were to retrain within the next two years.

Third, to elicit respondents' willingness to pay for further training, we ask them about the percentage of their wages that they would be willing to forgo while completing further training of at least 120 hours outside their company. To answer this question, respondents can indicate a number from zero to 100, where zero indicates that respondents would not be willing to forgo any part of their current wage.

5.3.4 Sample Balance and Descriptive Statistics

We perform a balancing test to check whether the randomization worked as intended, i.e., whether respondents' observable characteristics are balanced between treatment and control groups (see Appendix Table A5.2). Reassuringly, only two out of 28 comparisons in the

[15] Respondents are asked to what extent they agree on a five-point Likert scale with the following statements: *1) I am concerned about my professional future. 2) I will have different tasks in my job in the future than I have now. 3) I have a low risk of becoming unemployed. 4) I am concerned that new technologies will replace many tasks in my job. 5) I believe that my job will no longer exist in a few years. 6) I expect to be paid a higher wage in the future. 7) I will work on more demanding tasks in the future. 8) I will work fewer hours in the future than I do now because computers and computer-controlled machines will replace some of my activities. 9) I will work a lot with computers or computer-controlled machines in the future.* The index *labor-market concerns* combines the following items: being concerned about the professional future, having a low unemployment risk, being concerned that many job tasks will be replaced, being concerned that the occupation will no longer exist in the future, and expecting a higher wage. The index *work-environment change* comprises the other four items, i.e., expecting different tasks and more demanding tasks, increasingly working with machines and computers, and working fewer hours.

observable characteristics are statistically significant at the five percent level. In addition, item non-response is not correlated with treatment status (see Appendix Table A5.3).[16]

Respondents in our sample work in 1,118 different occupations. The most common occupations are “Management Assistant – Office Management”, “Office Clerk”, and “Bank Clerk”. This corresponds well to administrative data from the Federal Employment Agency demonstrating that most respondents in Germany work in the occupational group “office and secretariat” (Statistik der Bundesagentur für Arbeit, 2023b). The average automatability across all occupations in our sample is 51.3 percent. Occupations with the highest automatability (100 percent) are, for example, “Administrative employee” and “Machine, plant, and container cleaner”. Occupations with the lowest automatability (zero percent) are occupations such as “Social Worker/Social Pedagogue” and “Care Worker/Everyday Companion”.

We also match respondents’ reported occupations to additional data on the typical requirement level and the occupational field using the German Classification of Occupations from 2010 (Paulus and Matthes, 2013). This classification divides occupations into four requirement-level categories, distinguishing between unskilled or semi-skilled activities, specialist activities, complex specialist activities, and highly complex activities. We can compare the distribution across requirement levels with administrative data from the Federal Employment Agency (Statistik der Bundesagentur für Arbeit, 2023a). As it turns out, most respondents in our sample (in administrative data), 52.3 percent (57.3 percent), work in specialist activities, and only 6.1 percent (15.7 percent) in unskilled or semi-skilled activities. 22.9 percent (13.1 percent) work in complex specialist activities, while 18.7 percent (13.8 percent) work in highly complex activities. We, therefore, show a slight overrepresentation of respondents with high requirement levels of complex and highly complex activities, compared to unskilled or semi-skilled and specialist activities. As the automatability of occupations tends to be higher for lower-skilled occupations, our results could slightly underestimate effects for the general population.

In terms of occupational fields, our sample encompasses all major occupations pertaining to the German population (see Appendix Table A5.4). The majority of respondents work in administration and organization occupations (24.2 percent), whereas the smallest proportion works in agriculture (1.6 percent). Compared to the German population, workers in production and manufacturing are underrepresented in our sample (14.2 percent vs. 20.7 percent in the population). By contrast, workers in commercial services are slightly overrepresented in our sample (16.4 percent vs. 11.4 percent).

[16] In our preferred specification, we do not constrain respondents to having valid answers for all items of the survey. Due to item non-response, this means that observation numbers vary slightly across different specifications. Results remain virtually unchanged if we restrict the sample to respondents who answered all questions. Details available upon request.

5.4 Results

We first present descriptive evidence (section 5.4.1) on respondents' prior beliefs as well as correlations between prior beliefs, the occupation's automatability, and respondents' stated likelihood of participating in further training. Section 5.4.2 presents the experimental results. Section 5.4.3 discusses reasons for or against participating in further training.

5.4.1 Descriptive Results

Are Respondents Aware of Their Occupations' Automatability?
First, we examine whether respondents are aware of their occupations' automatability. Figure 5.1 shows the distribution of the automatability for the occupations of respondents in our sample calculated by experts from the IAB (transparent bars) and respondents' prior beliefs about their occupations' automatability (blue bars). Occupations' automatability is distributed fairly evenly from zero to 100 percent, while respondents' beliefs are skewed towards lower automatability.

We plot the correlation between respondents' beliefs and their occupation's automatability in Figure 5.2. Respondents on the 45-degree line (allowing for a five percentage point deviation above and below the actual value) state beliefs in line with their occupations' automatability. A majority of 67.5 percent of respondents underestimate their automatability, while 21.4 percent overestimate it.[17] The correlation between respondents' beliefs and their occupation's automatability is statistically significantly positive but very small (Figure 5.2). This implies that respondents in occupations with a high automatability have the largest misperceptions of their automatability. Appendix Figure A5.2 emphasizes that this is driven by differences in automatability, not differences in beliefs: both groups of respondents (with high and low automatability occupations) indicate similar beliefs of their occupations' automatability on average (blue shapes). Comparing the medians of actual shares and the beliefs between the two groups of respondents shows that the difference is only 5 percentage points for respondents with a low automatability (20.0 percent vs. 25.0 percent) while the difference is 45 percentage points for respondents with a high automatability (30.0 percent vs. 75.0 percent).

Overall, our results document sizeable misperceptions of automatability and profound optimism. This is in line with the literature documenting favorable misperceptions in other labor-market beliefs, for example, optimistic bias in job seekers' beliefs about their job-finding probability (Mueller et al., 2021) or underestimation of earnings at another potential employer (Jäger et al., 2021). However, on average, respondents in occupations with a higher automata-

[17] Note that rounding is an important concern in the elicitation of continuous beliefs in surveys (Manski and Molinari, 2010). We can assess the prevalence of rounding in our context by identifying respondents' rounding behavior throughout the questionnaire. In total, we use four questions in the experiment where respondents indicate continuous answers between zero and 100 with a slider. Only nine percent of respondents indicate a number which is a multiple of ten for each of the four sliders. This suggests that rounding is not of concern in our study.

bility are less confident about their answer (mean values 4.6 vs. 4.3 on a seven-point scale, difference statistically significant).

Which Respondents are Likely to Participate in Further Training?

In this section, we explore whether the likelihood of participating in further training and retraining correlates with the automatability of respondents' occupations or their beliefs. In particular, we investigate to what extent (individual beliefs about) the automatability of one's own occupation correlate with labor-market behavior.

In the control group, respondents state an average likelihood that they will participate in further training (retraining) of 40.7 percent (27.1 percent). The stated likelihood of participating in further training declines slightly with the automatability of respondents' occupations. Respondents in occupations with an automatability below 50 percent state on average a 44.1 percent likelihood of participating in further training, 6.5 percentage points statistically significantly higher than respondents in occupations with a high automatability.[18] Regarding retraining, respondents in occupations with a low automatability state a 28.8 percent likelihood of participating, on average, while respondents with a high automatability state a likelihood of 25.6 percent of participating (difference statistically significant at five percent level). This corroborates that the lower participation in further training for respondents with a high share of routine and automatable core tasks documented in the literature already emerges in stated participation expectations (Heß et al., 2019).

As documented in the previous section, a large proportion of respondents underestimates their occupation's automatability. It seems likely that individual training decisions are based on respondents' perceptions of how future skills will develop in their occupation. As it turns out, respondents' stated likelihood of participating in further training increases with beliefs about their occupations' automatability for an automatability below 50 percent and declines for higher levels (Figure 5.3).[19] Similarly, respondents with the lowest beliefs about their occupations' automatability state the lowest likelihood of participating in retraining. For this outcome, we see a positive relationship throughout: the higher the beliefs about the occupations' automatability, the higher the stated likelihood of participating in retraining. These results could suggest that respondents perceive further training as especially useful if their occupation's susceptibility to automation is moderate. In contrast, retraining seems to be

[18] This pattern is also reflected in respondent's previous participation in further training. Respondents with a high automatability of their occupation participated in fewer further training measures in the past than those with a low automatability (Appendix Figure A5.3). Furthermore, respondents in high-automatability occupations who did not previously participate in any training state the lowest likelihood of participating (32.1 percent) compared to those who previously participated (43.3 percent), while respondents in low automatability occupations with (without) previous experience in training state a 47.6 (34.8) percent likelihood of participating. These two differences are statistically significant. This indicates that respondents who might benefit the most have a lower likelihood of planning to participate in such measures.

[19] Figure 5.3 plots the unconditional relationships between respondents' beliefs and their stated likelihood of participating in further training (dark green) and retraining (light green). The results remain similar when controlling for age and gender.

an option seen as increasingly desirable the greater an occupation's exposure to technological advances.

Subsequently, we discuss results from our survey experiment where treated respondents receive information on their occupation's automatability. The experimental setup allows us to test whether information provision causally relates to changes in the stated likelihood of participation in further training.

5.4.2 Experimental Results

The Econometric Model

We estimate the effect of the information treatment on outcomes with the following regression model:

$$y_i = \alpha_0 + \alpha_1 T_i + \delta' X_i + \epsilon_i, \tag{5.1}$$

where y_i is the outcome variable of interest for respondent *i*, i.e., labor-market expectations or likelihood of participating in further training. T_i indicates whether respondent *i* received information on the automatability of his or her occupation. X_i is a vector of control variables, and ϵ_i is the error term. Since ϵ_i is uncorrelated with treatment status through randomization, the coefficient α_1 provides an unbiased estimate for the causal treatment effect of information provision even without adding further control variables. As the inclusion of control variables can increase the precision of estimates, we show results with control variables in our main analyses.

As previously discussed and supported by existing research, individuals in occupations with a high automatability are strongly underrepresented in continued education and further training initiatives (Heß et al., 2019; OECD, 2021) despite the fact that they would benefit most from further training. Since we hypothesize that the information provision has larger effects on outcomes for respondents in occupations with a high automatability (larger than 50 percent) because they hold biased beliefs about their occupations' automatability, we include an interaction term of the treatment indicator and a dummy variable *low occupation automatability* (LOA_i) in a second model:

$$y_i = \alpha_0 + \alpha_1 T_i + \alpha_3 LOA_i + \alpha_4 T_i \times LOA_i + \delta' X_i + \epsilon_i. \tag{5.2}$$

With this specification, we examine whether treatment effects differ for respondents in occupations with different automatabilities.[20]

[20] We would expect the size of the treatment effects to also correlate with the absolute distance between respondents' prior beliefs and the information treatment. In our sample, we do not have enough power to detect these effects although results show the expected pattern (results available on request).

We correct for multiple hypothesis testing for the outcomes pre-registered as primary outcomes, applying the correction proposed by List et al. (2019), whose procedure is based on Romano and Wolf (2010). Overall, correcting for multiple outcomes does not change the interpretation of our results. We report corrected p-values in the table notes of the respective regression tables of the outcomes that we pre-registered as primary outcomes (see Tables 5.2, 5.3, and 5.4).

Main Results

Beliefs About Own Job's Automatability. First, we examine to what extent information on the automatability of respondents' occupations is relevant to respondents' perceptions about their current job. As discussed in section 5.3.3, the task profile of employees within a given occupation could vary widely across individual jobs (Dengler and Matthes, 2018b). Similarly, respondents might take private information about the idiosyncrasies of their own job, or their current employer into account, leading them to believe that their own job's automatability might differ from the occupation-wide average. To this end, we estimate treatment effects of information provision on respondents' beliefs about the share of tasks that they specifically perform in their job that can be automated.

Table 5.1 shows that respondents in the treatment group on average increase their beliefs about the automatability of their own job when provided with information about their occupation's automatability (column 1). Since respondents, on average, underestimate the share of automatable tasks in their occupation (see section 5.4.1), this suggests that respondents update their beliefs in line with the information provided. On average, respondents in the control group believe that 26.8 percent of core tasks in their current jobs are automatable, which is significantly below the average automatability for their occupations calculated by the IAB (52.1 percent). On average, treated respondents state a 5.1 percentage point higher share of automatable core tasks (column 1). Furthermore, Appendix Figure A5.4 which shows the correlation between respondents' beliefs about their own job's automatability and their occupation's automatability separately for control and treatment groups, also shows this pattern. For the control group, beliefs about their own job's automatability are largely flat across the whole distribution. For the treatment group, we see that respondents' beliefs move upward if they receive information that their occupation's automatability is high. Column 3 of Table 5.1 reports treatment effects on the difference between the automatability of respondents' occupations and their beliefs about their own job's automatability. On average, the control group reports beliefs about the automatability of their own job that are 25.3 percentage points below the average automatability of occupations. The negative treatment effect of information provision indicates that the difference reduces by 7.0 percentage points in the treatment group.

We see how beliefs about respondents' own job automatability increase to a far greater extent for those respondents who receive information that their occupation's automatability is estimated to be above 50 percent. Columns 2 and 4 of Table 5.1 report results based on equation (5.2), reporting treatment effects separately for the subgroup of respondents in occupations with a high automatability. In the control group, respondents in the subgroup with high-automatability occupations believe on average that their own job's automatability is 30.3 percent, only 7.3 percentage points higher than respondents in occupations with a low automatability. In the treatment group, this share increases by 12.9 percentage points, while there is a small average reduction in beliefs for those in occupations with a low automatability by 2.4 percentage points (column 2). This treatment effect for respondents in high-automatability occupations corresponds to a 13.9 percentage point reduction in the difference between their occupations' automatability and their own stated beliefs about their job (column 4). This implies that respondents in occupations with a high automatability significantly update their beliefs on their own job's automatability when they receive the information treatment.

Labor-Market Expectations. Second, we report estimates showing that information on respondents' occupations' automatability increases concerns and changes expectations about their future work environment. Table 5.2 regresses the two indices of labor-market concerns and work-environment change (see section 5.3.3) on the treatment indicator, estimating equation (5.1); and on the treatment indicator, a dummy variable for respondents in an occupation with low automatability and their interaction estimating equation (5.2).[21]

The results in Table 5.2 show that respondents in the treatment group are more concerned about their professional future (column 1). The index for labor-market concerns increases by 9.7 percent of a standard deviation. This is primarily driven by those in occupations with a high automatability (column 2): the treatment effect for this group amounts to 15.9 percent of a standard deviation, while the treatment effect for those in occupations with a low automatability is small and insignificant.

Treated respondents are also more likely to expect changes in their work environment in the future (column 3). The index work-environment change increases in the treatment group by 13.0 percent of a standard deviation. For respondents in occupations with a high (low)

[21] In the control group, more than half of the respondents agree that they will have more demanding tasks in the future (53.6 percent), that they will work more with computers or machines (51.9 percent), that they have a low risk of unemployment (64.9 percent) and that they will earn a higher wage in the future (58.5 percent). Still, 39.4 percent of respondents agree that they will perform other tasks in the future. 30.1 percent of respondents agree that the are concerned about their professional future and 30.4 percent that they are concerned that many of their occupation tasks will be automated in the future. Only one quarter (25.3 percent) of respondents agrees that they will work fewer hours in the future because their tasks will be replaced and only 18.4 percent of respondents agree that their occupation might not exist in the future anymore (see Appendix Table A5.5). We also present treatment effects for the individual items of the indices in Appendix Table A5.5.

automatability, the treatment effect is 16.0 percent (10.5 percent) of a standard deviation, although the difference between groups is not statistically significant (column 4).[22]

Taken together, we conclude that information provision on the automatability of respondents' occupations has a significant effect on their labor-market expectations: treated respondents become more concerned about their professional future, and they more often agree that they will have different tasks and work fewer hours. This suggests that respondents seem to have a nuanced interpretation of the information on automatability, incorporating both labor-displacing and labor-reinforcing narratives of technological change.

Participation in Further Training and Retraining. Third, we show that information on the occupations' automatability increases the stated likelihood of participating in further training, and of participating in retraining. Table 5.3 reports average results and treatment effects. In the control group, respondents state a 40.7 percent likelihood of participating in further training on average, and a 27.1 percent likelihood of retraining. Treated respondents are 2.3 percentage points more likely to state that they will participate in further training (column 1) and 3.4 percentage points more likely to be willing to retrain (column 3). These effects correspond to 5.8 percent and 12.5 percent of the mean, respectively. Appendix Figure A5.5 reports the distribution of reported likelihoods of participation for the control group (blue bars) and the treatment group (transparent bars). For further training, we see that the information treatment tends to reduce the share of respondents who report a probability of participation close to zero (Panel A, not statistically significant). For participation in retraining, we observe a much greater share indicating a zero probability and a lower share indicating a 100 percent probability (Panel B). However, the information treatment significantly increases the share of respondents who report a probability of 100 percent to participate in retraining.

As discussed in section 5.4.1, the intended likelihood of participating in further training and retraining is slightly lower for respondents in occupations with a high automatability (see also baseline mean in Table 5.3). As it turns out, the effect of information provision is more pronounced for this group: treated respondents in occupations with a high automatability state a 4.6 percentage point greater likelihood of participating in further training and a 5.2 percentage point greater likelihood of retraining compared to uninformed respondents (see Table 5.3, columns 2 and 4).[23] The treatment effect for respondents in occupations with a low automatability is not statistically significant for either outcome. Thus, the information reduces the gap in willingness to participate in further training between those in high and

[22] As explained, we correct for multiple hypothesis testing for the primary outcomes. The average treatment effects on both indices remain statistically significant when we correct for multiple hypothesis testing in Table 5.2. They also remain statistically significant for respondents with a high share of automatable core tasks, but the interaction terms are no longer statistically significant.

[23] The two average treatment effects on the likelihood of participating in further training and retraining remain statistically significant when correcting for multiple hypothesis testing, same as the effects for high-automation participants. The interaction terms for both outcomes of participating in further training and retraining are not statistically significant when correcting for multiple hypothesis testing.

low automatability occupations by 95 percent and completely closes it for respondents' willingness to participate in retraining.

Willingness to Forgo Part of the Wage. Finally, we illustrate how information on the automatability of respondents' occupations increases the share of their wage that respondents are willing to forgo while participating in further training. Table 5.4 shows that information provision leads to an increase of 1.3 percentage points in the share of wages that respondents are willing to forgo to participate in further training (column 1). Again, this is driven by those in occupations with a high automatability. Treated respondents in occupations with a high automatability state a 2.7 percentage point greater share of their wage that they are willing to forgo (column 2) compared to treated respondents with a low automatability. The treatment effect for those in occupations with a high automatability is 2.6 percentage points. In addition, information provision slightly increases the share of those willing to forgo any positive share of their wage for respondents in occupations with a high automatability (column 4).[24]

Further Outcomes

Occupation Choice.[25] We also ask all respondents, if they were to retrain in the next two years, which occupation they would choose to retrain. This allows us to study the characteristics of aspirational occupations compared to the status quo.

On average, respondents in the treatment group report retraining occupations with a slightly higher automatability, although this effect is not statistically significant (see Table 5.5, column 1). We therefore find no evidence that respondents plan to retrain to occupations less exposed to technological change. One possible explanation for this result is that respondents continue to be uninformed about the automatability of occupations that are not their current one, even as they receive the information treatment. Given the large misperceptions of respondents' own occupation documented in section 5.4.1, it is plausible that respondents are unaware of the automatability of other occupations. We also do not find a statistically significant difference in treatment effects between high- and low-automatability respondents.

We also analyze whether respondents want to retrain to another occupational field and, more specifically, to an occupational field with a low automatability, and whether they want to retrain to higher requirement level occupations than their current occupation. Treated respondents are slightly more likely to indicate an occupation in a different occupational field (column 3). However, this is only true for respondents in occupations with low automatability (column 4). In addition, treated respondents are not more likely to indicate a retraining occupation in an occupational field with a low automatability (i.e., "health and social services",

[24] The average treatment effect as well as the treatment effect for respondents with a high share of automatable core tasks remain statistically significant when correcting for multiple hypothesis testing. The effects on the share of those willing to forgo any positive share of their wage become insignificant.
[25] While we pre-registered the question on which retraining occupations respondents would choose, analyses in this section were not pre-registered and are explorative.

“social science”, and “agriculture”) (columns 5 and 6). Information provision does not affect the requirement level of the retraining occupation (columns 7 and 8). However, treated respondents tend to state a lower-paying occupation (though not statistically significant on average but marginally significant for respondents in occupations with a high automatability; see columns 9 and 10).

Taken together, this analysis suggests that even though treated respondents are more willing to retrain, they do not plan to switch to occupations with a lower automatability or higher incomes. Thus, this finding corroborates that individuals are generally not aware of the automatability of occupations.

Policy Preferences. The extent of technological change and its implications for a large number of occupations have made programs to increase access to further training a current policy priority in many countries. In this section, we investigate to what extent information on the automatability of one’s occupation affects respondents’ preferences for different policies aimed at increasing participation in further training more broadly.

First, respondents were asked whether they favor or oppose the policy proposal that during one’s professional life, every person whose occupation is affected by structural change and digitization should be required to participate in further training. Appendix Table A5.6 shows the results: 62.5 percent in the control group (rather or strongly) support this proposal. However, information provision does not statistically significantly affect respondents’ agreement to this proposal, and we do not find heterogeneous effects by respondents’ occupations’ automatability (columns 1 and 2). We also ask respondents whether further training is effective for keeping pace with structural change. Approximately three-quarters of respondents (76.7 percent) in the control group agree with this statement, and information provision does not statistically significantly change agreement with this statement (see Appendix Table A5.6, columns 3 and 4).

Similarly, treated respondents are not more likely to think that the need for further training for (1) all employees or (2) employees with the same occupation as themselves will increase.[26] In the control group, 66.2 percent of respondents believe that the need for further training will (rather or strongly) increase for all employees (see Appendix Table A5.6, column 5). Focusing on employees in the same occupation as respondents themselves, the control group’s share is much lower: only 50.4 percent think that the need for further training will (rather or strongly) increase (column 7). Hence, respondents perceive a greater need for those in other occupations to participate in further training than for those in their occupation. This could suggest that they do not necessarily see those ‘like them’ in their occupation as affected by structural change as those in other occupations, which is in line with optimistic biases documented for other labor-market decisions (e.g., Mueller et al., 2021). This conforms to the observation that

[26] Questions are worded as follows: *What do you think, will the need for further training for the following groups of people increase, decrease, or remain unchanged in the future? (1) For all employees in Germany. (2) For people who are in the same occupation as me.*

53.8 percent of Germans think that there are more losers than winners because of structural change, while only 26.6 percent see themselves as a loser of structural change (Werner et al., 2022).

Overall, these results show that respondents have a positive view of further training as a means to keep up with structural change and support proposals to extend participation in further training; albeit slightly less for their own occupation. We also find that respondents do not update their preferences on further training policies when they receive information on the automatability of their occupation. The absence of treatment effects on policy preferences is in line with the literature documenting such patterns for a variety of domains (e.g., Kuziemko et al., 2015; Alesina et al., 2018).

5.4.3 Reasons for (Non-)Participation

Finally, we ask all respondents about their reasons for (not) wanting to participate in further training. Respondents are asked to what extent they agree with a set of statements relating to financial, time, and employer constraints (see Appendix Table A5.7). Previous literature has identified these as potential barriers to participating in further training (see e.g., van den Berg et al., 2019; Müller and Wenzelmann, 2020; Osiander and Stephan, 2020). Almost half of the respondents state that they face financial and employer constraints (45.4 and 45.0 percent, respectively). 35.2 percent state that they face time constraints due to caring for children and other relatives. Approximately 39.5 percent also feel that they do not want to participate in measures offered by the Federal Employment Agency. As expected, information provision does not affect the agreement with these constraints since these resource and personal constraints are not influenced by the information provided (see Appendix Table A5.7).

In contrast, we find no evidence that, on average, respondents feel insufficiently informed. Only 38.6 percent of respondents in the control group want to receive further information about further training possibilities and finance options at the end of the survey. Treated respondents are, on average, also not more likely to request additional information (see Appendix Table A5.8).[27]

Overall, 66.5 percent see themselves as being well-equipped for their future career (see Appendix Table A5.9, column 1). At the same time, almost half of respondents, 47.5 percent, perceive a great need to participate in further training (column 3). A similar fraction, 47.6 percent, report that they are unsure whether further training will pay off for them (column 5). On average, information provision does not affect agreement with these statements. While information provision tends to have opposing effects depending on their occupations' automatability, treatment effects are generally small and insignificant (columns 2, 4 and 6).

[27] We ask respondents the following question: *Would you like to receive more information about further training opportunities, funding, and providers in Germany?*

We also examine whether respondents' age might be a reason why some respondents are reluctant to participate in further training.[28] One might expect that older workers who are closer to their retirement age state less willingness to participate in further training and react less strongly to the information provided. For example, Innocenti and Golin (2022) find that older respondents are less likely to report that they are concerned about losing their job due to automation and less likely to retrain. Table 5.6 shows that this also applies in our setting: while information provision positively affects younger respondents' likelihood of participating in further training (2.6 percentage points), the treatment effect for respondents older than 60 corresponds to -0.2 percentage points and is not statistically significant. A likely reason is that younger workers face a longer time horizon during which technological transformation might affect their careers and during which they could recoup the returns of working in their new occupations, while older employees have only a few years left to work until they reach the legal retirement age (age 67 in Germany). Results are similar for respondents' likelihood to retrain, with a slightly larger treatment effect for respondents in occupations with a high automatability (column 2).

In addition, we examine whether there are differences according to respondents' educational attainment. Respondents with higher educational attainment could have a higher consumption value of learning and their private costs of learning might be lower. Therefore, we divide respondents into two groups: respondents with and without a university entrance qualification. The treatment effects for those without a university entrance qualification correspond to almost five percentage points for the likelihood of participating in further training and in retraining (Table 5.6, columns 3 and 4). The treatment effect on the likelihood of participating in further training for respondents with a university entrance qualification is close to zero and not statistically significant (see Table 5.6, column 3). This result is in line with our main results, which are driven by respondents with a high-automatability occupation: there is a negative correlation between respondents' educational attainment and automatability.

5.5 Conclusion

Technological and structural change in the labor market are increasing the demand for new skills in the workforce. Further training and retraining therefore emerge as key elements to bridge the gap between initial education and current developments. Yet, participation rates are especially low for individuals in occupations with a high automatability. We show that one potential reason for the lack of training participation are misperceptions regarding the automatability of tasks commonly performed in different occupations. On average, respondents underestimate the automatability of their occupation, especially those in occupations with a high automatability. We find that information provision about the automatability of a person's occupation affects labor-market expectations: respondents report more concern about their professional future and more often agree to perform other (more demanding) tasks. These

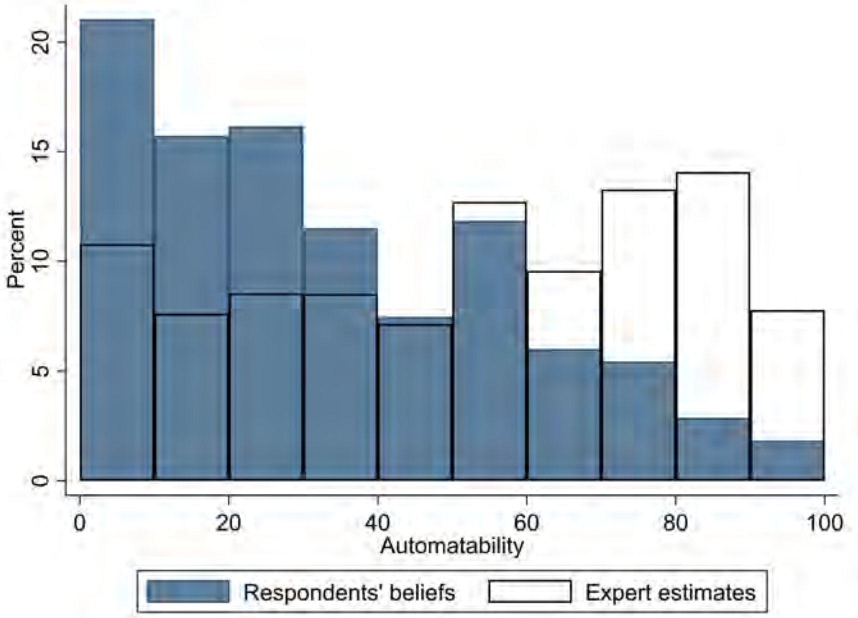
[28] We did not pre-register this heterogeneity.

results imply that respondents possess a nuanced understanding of the information regarding automatability, encompassing both narratives of technological change that displace labor and those that reinforce it. In addition, we find that providing information increases the willingness to participate in further training and retraining and increases the share of wages that respondents are willing to forgo to participate in further training. Treatment effects are larger for respondents in occupations with a high automatability and thus for those most likely at risk of being impacted by technological advances. Thus, providing information to respondents about their occupation's automatability can reduce the gap in intentions to participate in further training and even close the gap in intentions for retraining. Thereby, the inequality in training participation between individuals in high- and low-automatability occupations could be reduced.

Overall, our results show that providing the public with information about their occupations' automatability might help to raise awareness about the possible impact of technological change on the labor market, especially for those working in occupations at risk of automation. Thus, it might encourage people to participate in further training to prepare themselves for the changing requirements brought about by technological change. Further research is needed to explore whether the increased stated likelihood of participating in further training and retraining will translate into higher participation in further training in the real world and on a larger scale.

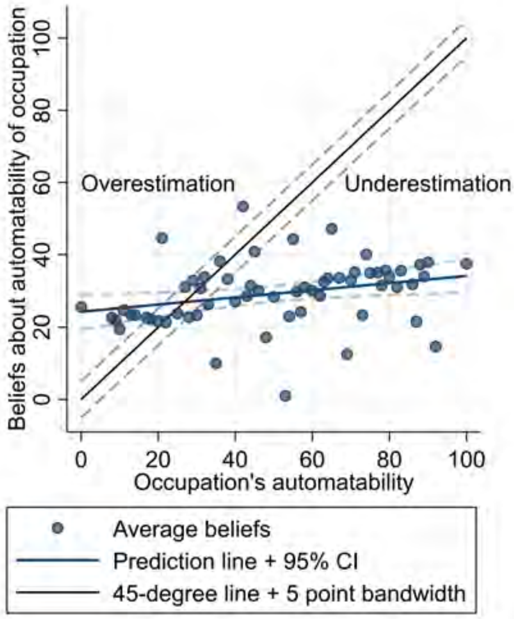
Figures and Tables

Figure 5.1: Distribution of Respondents' Beliefs and the Automatability of Respondents' Occupations



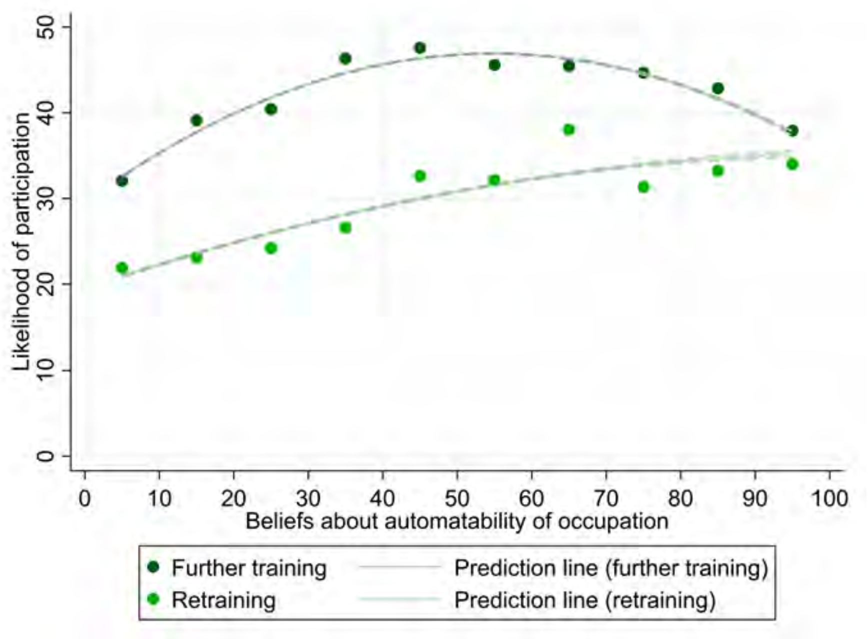
Notes: Blue bars depict bins for respondents' answers to the question *"What do you think is the percentage of core activities that people perform in the profession [answer from earlier question about current occupation] that can be automated?"*. Transparent bars depict bins for the automatability of respondents' occupations according to expert estimates by the Research Institute of the Federal Employment Agency in Germany (IAB). Data source: ifo Education Survey 2022.

Figure 5.2: Difference between Respondents' Beliefs and Occupations' Automatability



Notes: Respondents' answers to the question *"What do you think is the percentage of core activities that people perform in the profession [answer from earlier question about current occupation] that can be automated?"* are depicted as averages for each occupation's automatability (calculated by experts from the IAB). The blue line depicts the prediction line and the 95 percent confidence interval. The black line depicts the 45-degree line with a five-point bandwidth. Points above the 45-degree line indicate an overestimation of the occupation's automatability and points below indicate an underestimation. Data source: ifo Education Survey 2022.

Figure 5.3: Beliefs about Automatability and Likelihood of Participating in Further Training and Retraining



Notes: Respondents' answers to the questions *"How likely is it that you yourself will participate in further training of at least 120 hours within the next two years?"* and *"How likely is it that within the next two years, you will complete retraining to another occupation?"* are depicted as averages for 10 bins of respondents' beliefs about their occupation's automatability. The dark green dots depict the averages for the likelihood of participating in further training, and the light green dots depict the ones for retraining. Sample: control group respondents. Data source: ifo Education Survey 2022.

Table 5.1: Effect of Information Treatment on Beliefs

	Beliefs about own job's automatability		Difference occupation's automatability and beliefs about own job's automatability	
	(1)	(2)	(3)	(4)
Information	5.124***	12.874***	-6.989***	-13.858***
	(0.879)	(1.260)	(1.123)	(1.336)
Low automation		-7.299***		-43.694***
		(1.192)		(1.333)
Information x low automation		-15.321***		14.958***
		(1.646)		(1.801)
Covariates	Yes	Yes	Yes	Yes
Baseline mean	26.822	30.290	25.290	46.857
Observations	3,005	3,005	3,005	3,005
R-squared	0.056	0.168	0.063	0.394
Treatment effect for low automation		-2.447**		1.100
		(1.065)		(1.213)

Notes: OLS regressions. Dependent variables: (1) – (2) Beliefs about own job's automatability; (3) – (4) Occupation's automatability minus beliefs about own job's automatability. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. Covariates include: age, female, born in Germany, West Germany, living in large city, risk, patience, parents with university education, income, current employment status, middle school degree, high school degree, partner living in household, parental status, and imputation dummies. Outcomes in this table are pre-registered as secondary outcomes. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.2: Information Effect on Labor-Market Expectations (Index)

	Index labor-market concerns		Index work-environment change	
	(1)	(2)	(3)	(4)
Information	0.097***	0.159***	0.130***	0.160***
	(0.036)	(0.049)	(0.034)	(0.046)
Low automation		-0.194***		-0.129***
		(0.049)		(0.048)
Information x low automation		-0.118*		-0.055
		(0.071)		(0.067)
Covariates	Yes	Yes	Yes	Yes
Baseline mean	0.000	0.084	0.000	0.036
Observations	3,011	3,011	3,011	3,011
R-squared	0.088	0.103	0.148	0.155
Treatment effect for low automation		0.041		0.105**
		(0.050)		(0.049)

Notes: OLS regressions. Dependent variables: (1) – (2) Index labor-market concerns; (3) – (4) Index work-environment change. Randomized experimental treatment group “Information”: respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as primary outcomes. MHT corrected p-values: for “Information”: 0.032 in column (1), 0.015 in column (2), 0.001 in column (3), 0.001 in column (4). For “Information x low automation”: 0.159 in column (2), 0.414 in column (4). Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.3: Information Effect on Likelihood of Participating in Further Training and Retraining

	Likelihood further training		Likelihood retraining	
	(1)	(2)	(3)	(4)
Information	2.343**	4.647***	3.401***	5.194***
	(1.062)	(1.465)	(1.098)	(1.389)
Low automation		4.977***		1.860
		(1.498)		(1.359)
Information x low automation		-4.755**		-3.658*
		(2.132)		(2.007)
Covariates	No	Yes	No	Yes
Baseline mean	40.740	37.576	27.111	25.555
Observations	2,973	2,973	2,928	2,928
R-squared	0.142	0.145	0.188	0.189
Treatment effect for low automation		-0.108		1.535
		(1.541)		(1.441)

Notes: OLS regressions. Dependent variables: (1) – (2) Stated likelihood of participating in further training (between zero and 100); (3) – (4) Stated likelihood of participating in retraining (between zero and 100). Randomized experimental treatment group “Information”: respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as primary outcomes. MHT corrected p-values: for “Information”: 0.051 in column (1), 0.001 in column (2), 0.001 in column (3), 0.001 in column (4). For “Information x low automation”: 0.120 in column (2), 0.178 in column (4). Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.4: Information Effect on Willingness to Forgo Wage

	Willingness to forgo wage		Willingness to forgo wage > 0	
	(1)	(2)	(3)	(4)
Information	1.250**	2.612***	0.014	0.048*
	(0.614)	(0.871)	(0.017)	(0.024)
Low automation		0.354		0.026
		(0.841)		(0.025)
Information x low automation		-2.736**		-0.068*
		(1.225)		(0.035)
Covariates	Yes	Yes	Yes	Yes
Baseline mean	9.558	8.999	0.476	0.449
Observations	3,008	3,008	3,008	3,008
R-squared	0.103	0.105	0.106	0.107
Treatment effect for low automation		-0.124		-0.020
		(0.864)		(0.025)

Notes: OLS regressions. Dependent variables: (1) – (2) Share of wage willing to forgo (zero to 100); (3) – (4) Dummy variable coded one if stated willingness is non-zero (positive). Randomized experimental treatment group “Information”: respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as primary outcomes. MHT corrected p-values: for “Information”: 0.079 in column (1), 0.010 in column (2), 0.397 in column (3), 0.208 in column (4). For “Information x low automation”: 0.137 in column (2), 0.220 in column (4). Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.5: Characteristics of Retraining Occupations

	Automatability		Other occupational field		Low automation occupational field		Higher requirement level		Mean wage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Information	1.217	1.686	0.010	-0.028	-0.015	-0.035	-0.008	-0.031	-68.423	-117.155*
	(1.237)	(1.736)	(0.019)	(0.027)	(0.018)	(0.024)	(0.019)	(0.028)	(48.633)	(67.486)
Low automation		-9.578***		-0.089***		0.101***		-0.163***		-123.730*
		(1.728)		(0.028)		(0.026)		(0.027)		(68.479)
Information x low automation		-1.099		0.077**		0.040		0.046		98.639
		(2.440)		(0.039)		(0.037)		(0.038)		(97.380)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline mean	44.024	49.308	0.638	0.684	0.318	0.263	0.348	0.429	3,889.858	3,897.594
Observations	2,326	2,326	2,449	2,449	2,449	2,449	2,447	2,447	2,190	2,190
R-squared	0.027	0.054	0.010	0.014	0.049	0.066	0.017	0.039	0.146	0.147
Treatment effect for low automation		0.587		0.049*		0.005		0.015		-18.517
		(1.711)		(0.028)		(0.027)		(0.026)		(70.552)

Notes: OLS regressions. Dependent variables: (1) – (2) Automatability of the indicated retraining occupation (zero to 100); (3) – (4) Dummy variable indicating whether the retraining occupation is in another occupational field; (5) – (6) Dummy variable indicating whether the retraining occupation is in a low automation occupational field (health & social services, social science, and agriculture); (7) – (8) Dummy variable indicating whether the retraining occupation is of a higher requirement level; (9) – (10) Mean wage of the retraining occupation. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.6: Heterogeneity of Information Treatment Effects across Subgroups

	Age above 60		University entrance qualification	
	Likelihood further training	Likelihood retraining	Likelihood further training	Likelihood retraining
	(1)	(2)	(3)	(4)
Information [baseline: not in subgroup]	2.575**	3.683***	4.375***	4.667***
	(1.113)	(1.072)	(1.422)	(1.307)
Subgroup	-4.238	-4.838**	5.246***	-1.802
	(2.812)	(2.027)	(1.910)	(1.735)
Information x subgroup	-2.732	-3.378	-4.734**	-2.955
	(3.698)	(2.578)	(2.139)	(2.026)
Covariates	Yes	Yes	Yes	Yes
Baseline mean	42.288	28.932	35.349	24.548
Observations	2,973	2,928	2,970	2,925
R-squared	0.144	0.191	0.143	0.188
Treatment effect for subgroup	-0.157	0.305	-0.359	1.712
	(3.524)	(2.347)	(1.597)	(1.550)

Notes: OLS regressions. Dependent variables: (1) Stated likelihood of participating in further training (zero to 100); (2) Stated likelihood of participating in retraining (zero to 100); (3) Stated likelihood of participating in further training (zero to 100); (4) Stated likelihood of participating in retraining (zero to 100). Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Appendix A5.1 Institutional Background

This section provides an overview of the further training system in Germany. The system is characterized by a high degree of decentralization due to Germany's federal structure.

There are five types of learning provision: Basic Continuing Education and Training (CET), General CET, Vocational CET, CET in Higher Education and Adult Liberal Education (OECD, 2021). Basic CET refers to non-formal learning opportunities for adults lacking basic skills. General CET includes formal education opportunities for adults to obtain school leaving certificates. Vocational CET encompasses formal and non-formal learning opportunities covering different levels, ranging from basic vocational qualifications to Master crafts people, Bachelor's degrees and certified business economists (OECD, 2021). It also includes vocational retraining, adjustment training and vocational upskilling. CET in Higher Education includes Bachelor's and Master's degree programs, while Adult Liberal Education comprises learning opportunities offered by Adult Education Centers. In addition, there are several types of further training: in-company training, individual job-related, and non-job-related. In-company training is the most common type. The average duration for in-company further training is 29 hours, individual job-related approx. 153 hours, and non-job-related further training is 56 hours (Bundesministerium für Bildung und Forschung, 2019).

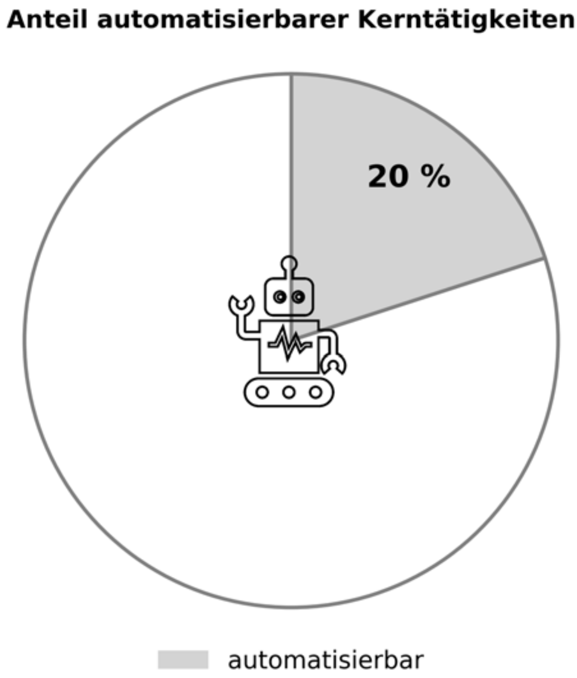
Overall, there are approximately 18,000 public and private further training providers (Bundesinstitut für Berufsbildung, 2020), including public institutions such as vocational schools or higher education institutions, CET institutions operated by enterprises, or groups of enterprises, social and economic partners such as trade unions and employer organizations, Chambers of Commerce and Trade, Chambers of Skilled Crafts, CET institutions run by churches, political parties, trade unions, foundations, other associations and Adult Education Centers.

The further training system in Germany is subject to various levels of regulation and is governed by numerous legal bases. These include collective bargaining agreements and company agreements, laws, and regulations at the state level. Companies, employees, and the public sector share the responsibility and obligation for further vocational training and its funding. According to the IW Continuing Education Survey 2020, the participation rate in further education by German companies was approximately 88 percent in the year 2019 (Seyda and Placke, 2020). Furthermore, the ifo Education Survey reported that 63 percent of respondents stated that they had participated in further training in the past. Conversely, this means that more than one third (37 percent) have not yet participated in any further training (Werner et al., 2022).

Furthermore, the German government provides support for further education through assistance and funding programs. For example, the Federal Employment Agency offers financial support to job seekers who wish to participate in further training measures to enhance their employability. Additionally, there are various government educational grants and tax benefits available to companies that invest in the further education of their employees.

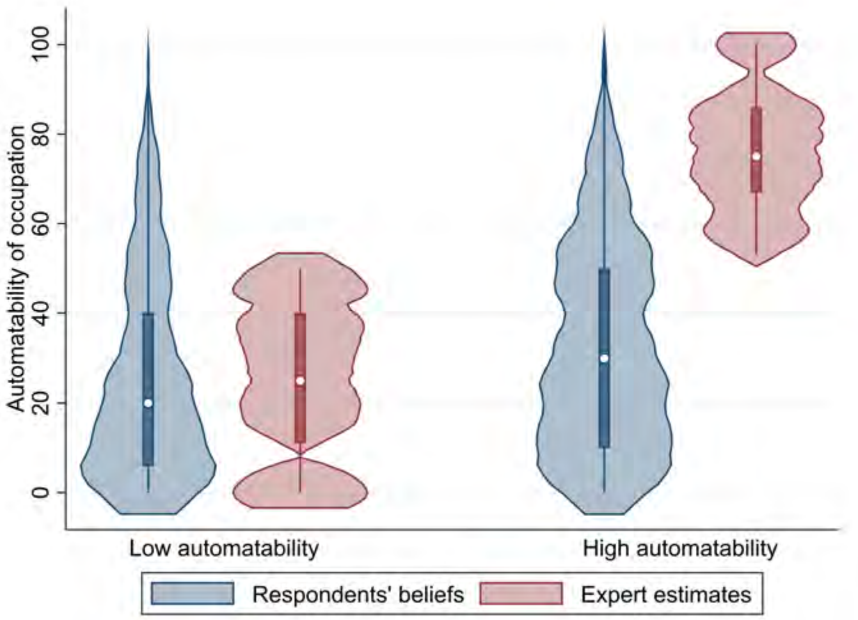
Appendix A5.2 Appendix Figures and Tables

Figure A5.1: Visual Representation of Information



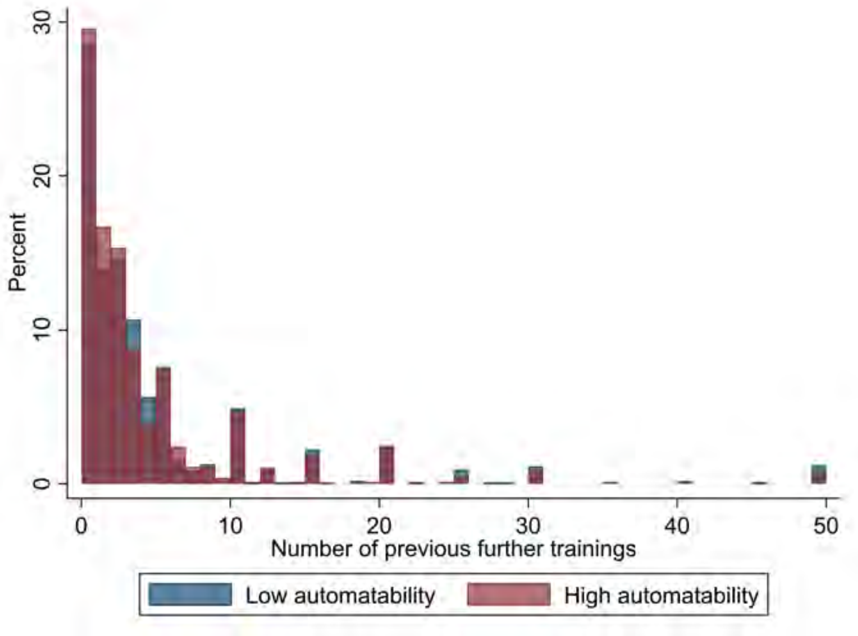
Notes: Example of the information about the occupation's automatability provided to the treatment group. This graph shows an example for an occupation with 20 percent automatable core tasks.

Figure A5.2: Distribution of Respondents' Beliefs and Occupations' Automatability for High- and Low-Automatable Occupations



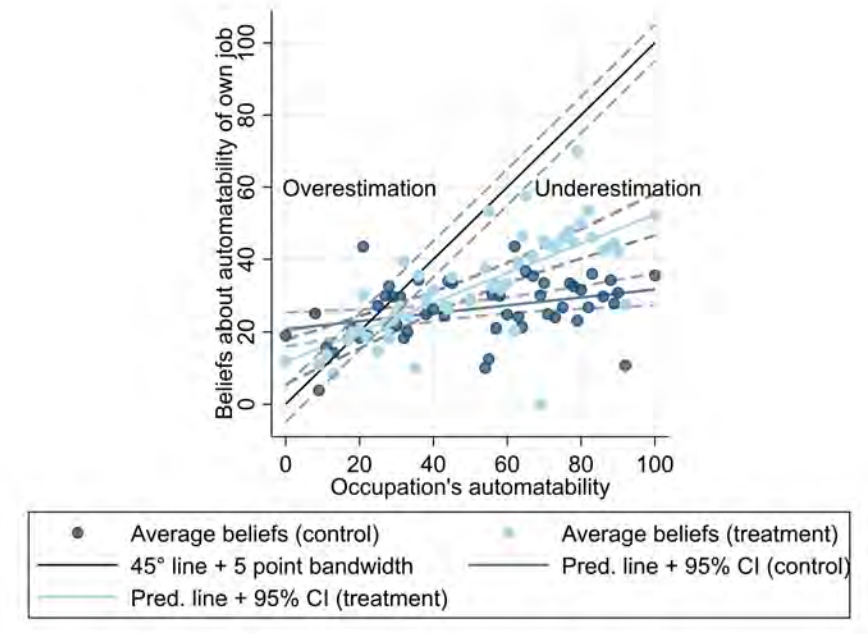
Notes: Blue shapes show the distribution of respondents' prior beliefs about their occupation's automatability, and red shapes show respondents' occupation's automatability according to expert estimates from the IAB, by respondents' occupations' automatability (high vs. low automatability). The box plots within the shapes show the median value and the interquartile range, with the extended lines representing upper- and lower-adjacent values. Data source: ifo Education Survey 2022.

Figure A5.3: Previous Participation in Further Training



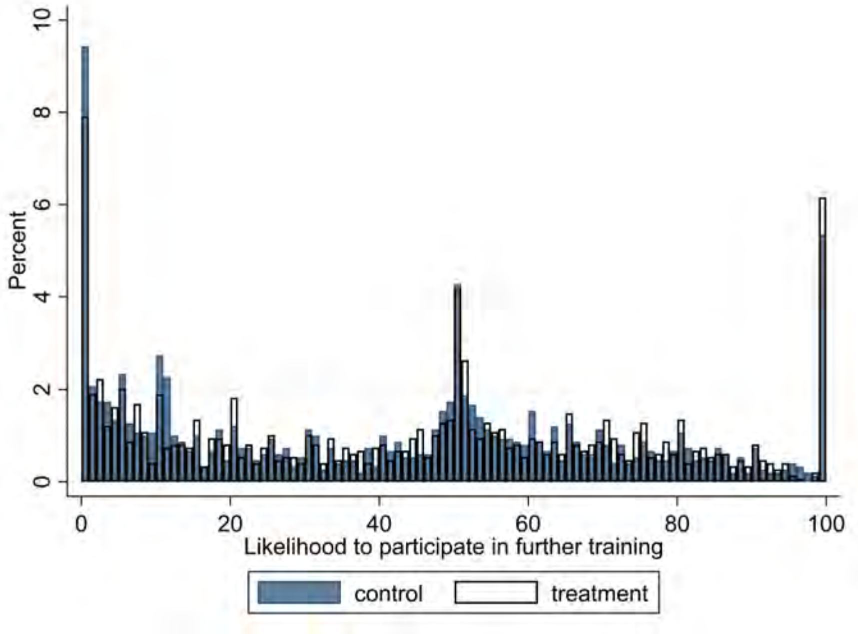
Notes: Previous participation in further training of respondents, divided into two groups: those in occupations with a high automatability (above 50 percent) and those in occupations with a low automatability (below 50 percent). Data source: ifo Education Survey 2022.

Figure A5.4: Respondents' Beliefs vs. Occupations' Automatability by Treatment Group

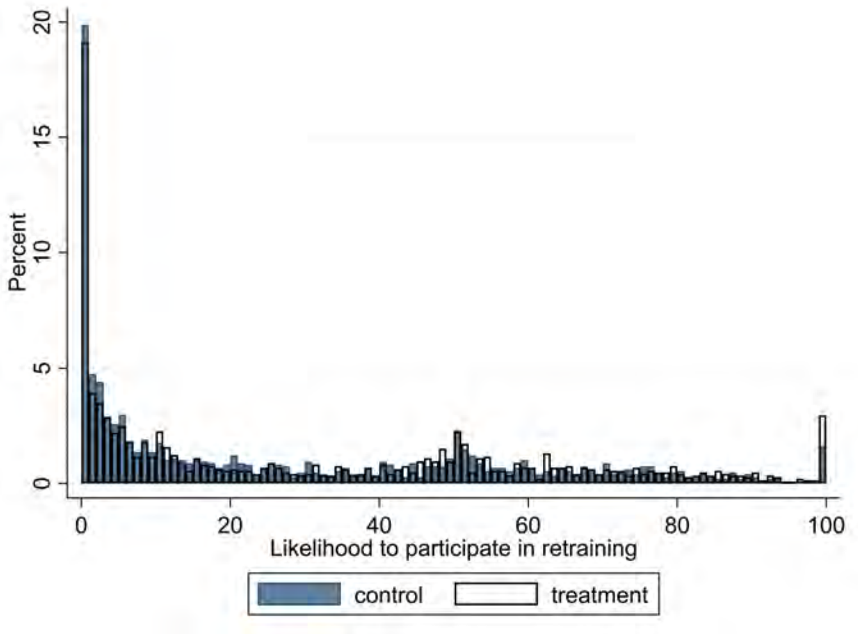


Notes: Respondents' answers to the question *"What do you think is the percentage of core activities that people perform in the profession [answer from earlier question about current occupation] that can be automated?"* are depicted as averages for each occupation's automatability (calculated by experts from the IAB), by respondents' treatment status. Randomized experimental treatment group: respondents informed about automatability of occupation. The black line depicts the 45-degree line with a five-point bandwidth. Points above the 45-degree line indicate an overestimation of the occupation's automatability and points below indicate an underestimation. Data source: ifo Education Survey 2022.

Figure A5.5: Likelihood of Participation in Further Training and Retraining



Panel A: Further Training



Panel B: Retraining

Notes: Respondents' answers to the questions *"How likely is it that you yourself will participate in further training of at least 120 hours within the next two years?"* and *"How likely is it that within the next two years you will complete retraining to another occupation?"* are depicted by respondents' treatment status. Randomized experimental treatment group: respondents informed about automatability of occupation. Data source: ifo Education Survey 2022.

Table A5.1: Experimental Setup

Control group	Treatment group
Current Occupation	
Guess automatability of occupation + certainty	
Guess automatability of current job	Info: Guess automatability of current job
Labor-market expectations	Info: **Labor-market expectations**
Future need for further training	Info: Future need for further training
Likelihood of participating in further training	Info: **Likelihood of participating in further training**
Likelihood of participating in retraining	Info: **Likelihood of participating in retraining**
Which job if retrain	Info: Which job if retrain
Policy proposal: Obligation further training	Info: Policy proposal: Obligation further training
Willingness to forgo wage	Info: **Willingness to forgo wage**
Information acquisition	Info: Information acquisition
Barriers to participation in training activities	

Notes: Experimental setup. Questions in bold are **primary outcomes** as specified in the AEA Registry AEARCTR-0009464.

Table A5.2: Sample Balance

Control Variable	Control Mean	Treatment Mean	Difference	p-value
Age	42.65	42.98	0.33	0.49
Female	0.47	0.46	-0.01	0.67
Born in Germany	0.93	0.92	-0.02	0.12
City size $\geq$ 100,000	0.37	0.40	0.03	0.11
Partner	0.64	0.65	0.02	0.35
Parent(s) w/ university degree	0.38	0.38	-0.01	0.77
Highest educational degree				
No degree	0.24	0.26	0.02	0.16
Middle school degree	0.32	0.32	0.00	0.88
Univ. entrance degree	0.44	0.42	-0.02	0.28
Employment status				
Full-time	0.66	0.66	0.00	0.95
Part-time	0.22	0.23	0.01	0.50
Self-employed	0.07	0.05	-0.02	0.05
Unemployed	0.00	0.00	0.00	0.98
Retired/Ill/etc.	0.05	0.06	0.01	0.50
Parent status	0.54	0.54	-0.01	0.77
Party preference				
CDU/CSU	0.17	0.17	-0.01	0.61
SPD	0.18	0.17	-0.01	0.61
Grüne	0.13	0.12	-0.01	0.55
Linke	0.05	0.06	0.00	0.85
FDP	0.08	0.09	0.00	0.73
AfD	0.09	0.11	0.02	0.03
Other	0.02	0.02	0.00	0.75
None	0.27	0.27	-0.01	0.57
General voting	0.79	0.80	0.01	0.45
Patience	7.04	7.02	-0.02	0.84
Risk	5.72	5.78	0.06	0.51
Monthly household income (in EUR)	2944.21	3014.03	69.82	0.26
West Germany	0.81	0.80	-0.01	0.56

Notes: Group means. “Difference” displays the difference in means between the control group and the treatment group who received the information about the occupation’s automatability. Data source: ifo Education Survey 2022.

Table A5.3: Item Non-Response

Control Variable	Control Mean	Treatment Mean	Difference	p-value
Labor-market expectation				
Concerned future	0.001	0.002	0.001	0.311
Other job tasks	0.000	0.001	0.001	0.157
Low risk unemployment	0.001	0.001	0.001	0.555
Automation tasks	0.000	0.001	0.001	0.317
Occupation existence	0.000	0.001	0.001	0.317
Higher wages	0.000	0.001	0.001	0.317
More demanding tasks	0.000	0.001	0.001	0.317
Less hours	0.000	0.001	0.001	0.157
More computer tasks	0.000	0.001	0.001	0.157
Need further training: all employees	0.001	0.000	-0.001	0.317
Need further training: same job employees	0.001	0.000	-0.001	0.317
Likelihood further training	0.015	0.011	-0.004	0.280
Likelihood retraining	0.029	0.027	-0.002	0.712
Policy proposal: obligation further training	0.002	0.000	-0.002	0.083
Forgo wage	0.001	0.002	0.001	0.311
Information acquisition	0.000	0.001	0.001	0.157
Financial constraints	0.001	0.001	0.001	0.555
Time constraints	0.000	0.001	0.001	0.157
Employer constraints	0.000	0.001	0.001	0.317
FEA offered further training	0.000	0.001	0.001	0.317
Gains insecure	0.000	0.001	0.001	0.317
Great necessity	0.001	0.001	0.000	0.991
Confidence job future	0.000	0.001	0.001	0.317
Good measure structural change	0.001	0.001	0.000	0.991

Notes: Group means of item non-response. “Difference” displays the difference in means between the control group and the treatment group who received the information about the occupation’s automatability. Data source: ifo Education Survey 2022.

Table A5.4: Occupational Field of Respondents and German Population

Occupational Field	Share of Respondents in Respective Field	Share of German Population in Respective Field	Difference
Agriculture	1.59	1.50	0.09
Production & Manufacturing	14.18	20.71	-6.53
Construction & Architecture	4.02	6.10	-2.08
Natural Sciences	6.91	4.29	2.62
Transport	12.22	13.41	-1.19
Commercial Services	16.44	11.38	3.03
Administration & Organization	24.18	20.39	3.79
Health & Social Services	15.84	18.90	-3.06
Social Sciences	4.62	2.78	1.84

Notes: Shares of respondents in respective fields in respondent sample and in German population. Data sources: ifo Education Survey 2022 and Statistik der Bundesagentur für Arbeit 2022.

Table A5.5: Information Effect on Labor-Market Expectations

	Index LM concerns					Index work-environment change			
	Concerned future	Low risk un-employment	Automation tasks	Occupation existence	Higher wages	Other job tasks	More demanding tasks	Less hours	More computer tasks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Information	0.118*	-0.006	0.254***	0.224***	0.022	0.246***	0.067	0.177***	0.195***
	(0.068)	(0.068)	(0.067)	(0.065)	(0.064)	(0.065)	(0.060)	(0.065)	(0.069)
Low automation	-0.069	-0.235***	-0.307***	-0.112*	0.158**	-0.128*	-0.085	-0.151**	-0.360***
	(0.069)	(0.069)	(0.066)	(0.064)	(0.065)	(0.066)	(0.062)	(0.065)	(0.071)
Information x low automation	-0.091	0.025	-0.180*	-0.196**	0.055	-0.160*	0.029	-0.090	-0.078
	(0.098)	(0.096)	(0.094)	(0.092)	(0.090)	(0.094)	(0.087)	(0.093)	(0.100)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline mean	2.438	2.542	2.563	2.001	3.269	2.814	3.206	2.291	3.231
Control share rather/fully agree	0.301	0.649	0.304	0.184	0.585	0.394	0.536	0.253	0.520
Observations	3,008	3,009	3,011	3,011	3,011	3,010	3,011	3,010	3,010
R-squared	0.076	0.045	0.086	0.071	0.078	0.089	0.104	0.091	0.096
TE for low automation	0.027	0.019	0.074	0.028	0.077	0.086	0.096	0.087	0.117
	(0.070)	(0.068)	(0.066)	(0.065)	(0.064)	(0.068)	(0.063)	(0.066)	(0.072)

Notes: OLS regressions. Dependent variables (five-point scale: 1 = fully disagree, 5 = fully agree): (1) I am concerned about my professional future; (2) I have a low risk of becoming unemployed; (3) I am concerned that many tasks in my occupation will be replaced by new technologies; (4) I believe that my occupation will no longer exist in a few years; (5) I expect to be paid a higher wage in the future; (6) I will have different tasks in my occupation in the future than I have now; (7) I will work on more demanding tasks in the future; (8) I will work fewer hours in the future than I do now because some of my activities will be replaced by computers and computer-controlled machines; (9) In the future, I will work a lot with computers and computer-controlled machines. Randomized experimental treatment group “Information”: respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. Control share rather/fully agree: dummy variable indicating the share of respondents rather or fully agreeing to the statement. See Table 5.1 for included covariates. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.6: Information Effect on Policy Preferences

	Support compulsory further training		Further training is good strategy to cope with structural change		Need for further training will incease (all employees)		Need for further training will increase (same occupation)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information	0.011	0.015	0.014	0.009	-0.015	-0.009	0.026	0.031
	(0.017)	(0.024)	(0.015)	(0.021)	(0.017)	(0.024)	(0.018)	(0.025)
Low automation		0.035		0.015		-0.002		-0.037
		(0.024)		(0.022)		(0.024)		(0.025)
Information x low automation		-0.010		0.010		-0.011		-0.010
		(0.034)		(0.030)		(0.034)		(0.036)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline mean	0.625	0.605	0.767	0.760	0.662	0.654	0.504	0.514
Observations	3,009	3,009	3,010	3,010	3,011	3,011	3,011	3,011
R-squared	0.059	0.060	0.054	0.054	0.053	0.054	0.051	0.053
TE for low automation		0.005		0.019		-0.020		0.021
		(0.024)		(0.021)		(0.024)		(0.026)

Notes: OLS regressions. Dependent variables: (1) – (2) Policy proposal: obligation for all employees affected by structural change and digitization to participate in further training (dummy coded one if agree); (3) – (4) Further training is a good way to keep pace with structural change (dummy coded one if agree); (5) – (6) Need for further training for all employees (dummy coded one if increase); (7) – (8) Need for further training for employees in same occupation as oneself (dummy coded one if increase); Using the full variation of the five-point Likert scale does not change the interpretation of the results. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as secondary outcomes. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.7: Barriers to Training Participation (I)

	Financial constraints		Time constraints		Employer constraints		Offered by FEA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information	0.014	0.035	0.003	0.020	-0.013	-0.018	-0.001	0.004
	(0.018)	(0.025)	(0.017)	(0.024)	(0.018)	(0.025)	(0.018)	(0.025)
Low automation		0.003		0.008		-0.039		0.008
		(0.025)		(0.024)		(0.025)		(0.025)
Information x low automation		-0.041		-0.035		0.011		-0.010
		(0.035)		(0.034)		(0.035)		(0.035)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	0.454	0.457	0.352	0.345	0.450	0.473	0.395	0.390
Observations	3,009	3,009	3,010	3,010	3,011	3,011	3,011	3,011
R-squared	0.068	0.068	0.046	0.046	0.074	0.075	0.026	0.026
Treatment effect for low automation		-0.006		-0.015		-0.007		-0.006
		(0.025)		(0.024)		(0.025)		(0.025)

Notes: OLS regressions. Dependent variables (dummy = 1 if person fully or somewhat agrees): (1) – (2) I cannot financially afford to attend further training; (3) – (4) I do not have time for further training (e.g., because of caring for relatives, childcare, etc.); (5) – (6) My employer does not offer me the opportunity for further training; (7) – (8) I do not wish to participate in any further training funded by the Federal Employment Agency. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as secondary outcomes. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.8: Information Effect on Information Acquisition

	Information acquisition	
	(1)	(2)
Information	0.002	0.029
	(0.017)	(0.024)
Low automation		0.059**
		(0.025)
Information x low automation		-0.056
		(0.035)
Covariates	Yes	Yes
Baseline mean	0.386	0.350
Observations	3,010	3,010
R-squared	0.066	0.068
Treatment effect for low automation		-0.027
		(0.025)

Notes: OLS regressions. Dependent variables: (1) – (2) Information acquisition. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as secondary outcomes. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.9: Barriers to Training Participation (II)

	Confident about professional future		Great necessity to participate in further training		Unsure about returns to training	
	(1)	(2)	(3)	(4)	(5)	(6)
Information	-0.007	-0.020	-0.005	0.026	0.010	0.032
	(0.017)	(0.024)	(0.018)	(0.025)	(0.018)	(0.026)
Low automation		0.014		0.069***		-0.000
		(0.024)		(0.025)		(0.026)
Information x low automation		0.027		-0.063*		-0.044
		(0.034)		(0.035)		(0.036)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Baseline mean	0.665	0.658	0.475	0.435	0.476	0.481
Observations	3,011	3,011	3,010	3,010	3,011	3,011
R-squared	0.052	0.053	0.069	0.071	0.019	0.020
Treatment effect for low automation		0.007		-0.037		-0.012
		(0.024)		(0.026)		(0.026)

Notes: OLS regressions. Dependent variables (dummy = 1 if person fully or somewhat agrees): (1) – (2) I am well-equipped for my future; (3) – (4) I see a great need to participate in further training; (5) – (6) I am unsure whether further training will pay off for me. Randomized experimental treatment group "Information": respondents informed about automatability of occupation. Baseline mean: mean of the variable in the control group or baseline group. See Table 5.1 for included covariates. Outcomes in this table are pre-registered as secondary outcomes. Data source: ifo Education Survey 2022. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Acemoglu, Daron (2009). *Introduction to Modern Economic Growth*. Princeton: Princeton University Press.

Acemoglu, Daron and Pascual Restrepo (2018). "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment". *American Economic Review* 108(6), 1488–1542.

– (2019). "Automation and New Tasks: How Technology Displaces and Reinstates Labor". *Journal of Economic Perspectives* 33(2), 3–30.

– (2020a). "Robots and Jobs: Evidence from US Labor Markets". *Journal of Political Economy* 128(6).

– (2020b). "Unpacking Skill Bias: Automation and New Tasks". *AEA Papers and Proceedings* 110, 356–361.

– (2022). "Tasks, Automation, and the Rise in U.S. Wage Inequality". *Econometrica* 90(5), 1973–2016.

Aggarwal, Charu C. (2018). *Neural Networks and Deep Learning*. Cham: Springer International Publishing.

Aghion, Philippe and Peter Howitt (1992). "A Model of Growth Through Creative Destruction". *Econometrica* 60(2), 323–351.

Ahrens, Achim, Christian B. Hansen, and Mark E. Schaffer (2019). "lassopack: Model Selection and Prediction with Regularized Regression in Stata". *The Stata Journal* 20(1), 176–235.

Alesina, Alberto and George-Maria Angeletos (2005). "Fairness and Redistribution". *American Economic Review* 95(4), 960–980.

Alesina, Alberto and Paola Giuliano (2011). "Preferences for Redistribution". *Handbook of Social Economics*. Ed. by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson. Vol. 1. Elsevier. Chap. 4, 93–131.

Alesina, Alberto and Edward Glaeser (2004). *Fighting Poverty in the US and Europe*. Oxford: Oxford University Press.

Alesina, Alberto and Eliana La Ferrara (2005). "Preferences for Redistribution in the Land of Opportunities". *Journal of Public Economics* 89(5-6), 897–931.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). "Intergenerational Mobility and Preferences for Redistribution". *American Economic Review* 108(2), 521–554.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden (2020). "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* 128(5), 1753–1788.

Andersen, Simon Calmar, Maria Knoth Humlum, and Anne Brink Nandrup (2016). “Increasing Instruction Time in School Does Increase Learning”. *Proceedings of the National Academy of Sciences of the United States of America* 113(27), 7481–7484.

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

Angrist, Joshua D. and Jonathan Guryan (2008). “Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements”. *Economics of Education Review* 27(5), 483–503.

Angrist, Joshua D. and Victor Lavy (2001). “Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools”. *Journal of Labor Economics* 19(2), 343–369.

Angrist, Joshua D., Victor Lavy, Jetson Leder-Luis, and Adi Shany (2019). “Maimonides’ Rule Redux”. *American Economic Review: Insights* 1(3), 309–324.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton: Princeton University Press.

Arntz, Melanie, Sebastian Blesse, and Philipp Doerrenberg (2022). “The End of Work is Near, Isn’t It? Survey Evidence on Automation Angst”. *ZEW - Centre for European Economic Research Discussion Paper 22-036*.

Arntz, Melanie, Terry Gregory, and Ulrich Zierahn (2016). “The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis”. *OECD Social, Employment and Migration Working Papers 189*.

– (2017). “Revisiting the Risk of Automation”. *Economics Letters* 159, 157–160.

– (2020). “Digitalisierung und die Zukunft der Arbeit”. *Wirtschaftsdienst* 100(S1), 41–47.

Ash, Elliott, Sergio Galletta, and Tommaso Giommoni (2022). “A Machine Learning Approach to Analyze and Support Anti-Corruption Policy”. *Available at SSRN: 3589545*.

Ashenfelter, Orley and John Ham (1979). “Education, Unemployment, and Earnings”. *Journal of Political Economy* 87(5), 99–116.

Athey, Susan and Guido W. Imbens (2016). “Recursive Partitioning for Heterogeneous Causal Effects”. *Proceedings of the National Academy of Sciences of the United States of America* 113(27), 7353–7360.

– (2019). “Machine Learning Methods That Economists Should Know About”. *Annual Review of Economics* 11(1).

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). “Generalized Random Forests”. *The Annals of Statistics* 47(2).

Athey, Susan, Stefan Wager, Vitor Hadad, Sylvia Klosin, Nikolaj Muhelbach, Xinkun Nie, and Matt Schaelling (2020). “Estimation of Heterogeneous Treatment Effects - Prepared for “Machine Learning and Causal Inference” class”. *Available online: https://gsbdbi*

.github.io/ml_tutorial/hte_tutorial/hte_tutorial.html, *last accessed: 12/09/2023*.

Aucejo, Esteban M. and Teresa Foy Romano (2016). “Assessing the Effect of School Days and Absences on Test Score Performance”. *Economics of Education Review* 55, 70–87.

Autor, David H. (2022). “The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty”. *NBER Working Paper 30074*.

Autor, David H., Caroline Chin, Anna Salomons, and Bryan Seegmiller (2022). “New Frontiers: The Origins and Content of New Work, 1940–2018”. *NBER Working Paper 30389*.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger (1998). “Computing Inequality: Have Computers Changed the Labor Market?” *The Quarterly Journal of Economics* 113(4), 1169–1213.

Autor, David H. and Anna Salomons (2018). “Is Automation Labor Share-Displacing? Productivity Growth, Employment, and the Labor Share”. *Brookings Papers on Economic Activity* 2018(1), 1–87.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber (2011). *Multivariate Analysemethoden*. 13th ed. Berlin Heidelberg: Springer.

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein (2018). “Improving Refugee Integration Through Data-Driven Algorithmic Assignment”. *Science* 359(6373), 325–329.

Barrios-Fernández, Andrés and Giulia Bovini (2021). “It’s Time to Learn: School Institutions and Returns to Instruction Time”. *Economics of Education Review* 80, 102068.

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L. Miller, and Eleonora Patacchini (2023). “Refining Public Policies with Machine Learning: The Case of Tax Auditing”. *NBER Working Paper 30777*.

Battistin, Erich and Elena Claudia Meroni (2016). “Should We Increase Instruction Time in Low Achieving Schools? Evidence from Southern Italy”. *Economics of Education Review* 55, 39–56.

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

– (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. New York: National Bureau of Economic Research.

Bellei, Cristián (2009). “Does Lengthening the School Day Increase Students’ Academic Achievement? Results from a Natural Experiment in Chile”. *Economics of Education Review* 28(5), 629–640.

Bellemare, Charles, Sabine Kröger, and Arthur van Soest (2008). “Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities”. *Econometrica* 76(4), 815–839.

Bellenberg, Gabriele (2012). *Schulformwechsel in Deutschland: Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I*. Gütersloh: Bertelsmann Stiftung.

Bénabou, Roland and Efe A. Ok (2001). “Social Mobility and the Demand for Redistribution: The POUM Hypothesis”. *The Quarterly Journal of Economics* 116(2), 447–487.

Bénabou, Roland and Jean Tirole (2006). “Belief in a Just World and Redistributive Politics”. *The Quarterly Journal of Economics* 121(2), 699–746.

Berens, Johannes, Julian Burghoff, Kerstin Schneider, Simon Görtz, and Simon Oster (2019). “Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods”. *Journal of Educational Data Mining* 3(11).

Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann (2021). “Testing”. *Journal of Human Resources*, 0520–10886R1.

Bertrand, Marianne, Bruno Crépon, Alicia Marguerie, and Patrick Premand (2017). “Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d’Ivoire”. *Mimeo*.

Bessen, James (2019). “Automation and Jobs: When Technology Boosts Employment”. *Economic Policy* 34, 589–626.

Bessen, James, Maarten Goos, Anna Salomons, and Wiljan van den Berge (2020). “Automation: A Guide for Policymakers”. *Economic Studies at Brookings*.

Bietenbeck, Jan (2014). “Teaching Practices and Cognitive Skills”. *Labour Economics* 30, 143–153.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold (2018). “Africa’s Skill Tragedy”. *Journal of Human Resources* 53(3), 553–578.

Bingley, Paul, Eskil Heinesen, Karl Fritjof Krassel, and Nicolai Kristensen (2018). “The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement”. *IZA Discussion Paper 11807*.

Björklund, Anders and Kjell G. Salvanes (2011). “Education and Family Background: Mechanisms and Policies”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann. Vol. 3. Elsevier. Chap. 3, 201–247.

Blanas, Sotiris, Gino Gancia, and Lee, Sang Yoon (2019). “Who is Afraid of Machines?” *Economic Policy* 34(100), 627–690.

Blömeke, Sigrid, Rolf Vegar Olsen, and Ute Suhl (2016). “Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality”. *Teacher Quality, Instructional Quality and Student Outcomes*. Ed. by Trude Nilsen and Jan-Eric Gustafsson. Springer. Chap. 2, 21–50.

Blossfeld, Hans-Peter and Hans-Günther Roßbach, eds. (2011). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer Fachmedien.

Breiman, Leo (2001). “Random Forests”. *Machine Learning* 45, 5–32.

Breiman, Leo, Adele Cutler, Andy Liaw, and Matthew Wiener (2018). “Package “random forest”: Breiman and Cutler’s Random Forests for Classification and Regression”. *CRAN*.

Bublitz, Elisabeth (2022). “Misperceptions of Income Distributions: Cross-Country Evidence from a Randomized Survey Experiment”. *Socio-Economic Review* 20(2), 435–462.

Buchhorn, Jascha and Berthold U. Wigger (2021). “Predicting Student Dropout: A Replication Study Based on Neural Networks”. *CESifo Working Paper 9300*.

Buckley, Edward and Rachel Croson (2006). “Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods”. *Journal of Public Economics* 90(4-5), 935–955.

Buddin, Richard and Gema Zamarro (2009). “Teacher Qualifications and Student Achievement in Urban Elementary Schools”. *Journal of Urban Economics* 66(2), 103–115.

Bundesinstitut für Berufsbildung, ed. (2020). *Datenreport zum Berufsbildungsbericht 2020. Informationen und Analysen zur Entwicklung der beruflichen Bildung*. Bonn: Bundesinstitut für Berufsbildung.

Bundesministerium für Arbeit und Soziales, ed. (2021). *Forschungsbericht 526/3: Aktualisierte BMAS-Prognose “Digitalisierte Arbeitswelt”*. Berlin: BMAS.

Bundesministerium für Bildung und Forschung, ed. (2019). *Weiterbildungsverhalten in Deutschland 2018: Ergebnisse des Adult Education Survey - AES-Trendbericht*. Bonn: BMBF.

Bursztyn, Leonardo (2016). “Poverty and the Political Economy of Public Education Spending: Evidence from Brazil”. *Journal of the European Economic Association* 14(5), 1101–1128.

Cappelen, Alexander W., Astri Dange Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). “The Pluralism of Fairness Ideals: An Experimental Approach”. *American Economic Review* 97(3), 818–827.

Cappelen, Alexander W., Karl O. Moene, Erik Ø. Sørensen, and Bertil Tungodden (2013). “Needs Versus Entitlements - An International Fairness Experiment”. *Journal of the European Economic Association* 11(3), 574–598.

Cappelen, Alexander W., Knut Nygaard, Erik Ø. Sørensen, and Bertil Tungodden (2015). “Social Preferences in the Lab: A Comparison of Students and a Representative Population”. *The Scandinavian Journal of Economics* 117(4), 1306–1326.

Card, David (1999). “The Causal Effect of Education on Earnings”. *Handbook of Labor Economics*. Ed. by Orley C. Ashenfelter and David Card. Vol. 3A. Elsevier. Chap. 30, 1801–1863.

Card, David and Alan B. Krueger (1992). “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States”. *Journal of Political Economy* 100(1), 1–40.

Carroll, John B. (1989). “The Carroll Model: A 25-Year Retrospective and Prospective View”. *Educational Researcher* 18(1), 26–31.

Cattaneo, Maria A., Chantal Oggenfuss, and Stefan C. Wolter (2017). “The More, The Better? The Impact of Instructional Time on Student Performance”. *Education Economics* 25(5), 433–445.

Chan, Kenneth S., Stuart Mestelman, Rob Moir, and R. Andrew Muller (1996). “The Voluntary Provision of Public Goods under Varying Income Distributions”. *The Canadian Journal of Economics* 29(1), 54–69.

Chandler, Dana, Steven D. Levitt, and John A. List (2011). “Predicting and Preventing Shootings among At-Risk Youth”. *American Economic Review* 101(3), 288–292.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates". *American Economic Review* 104(9), 2593–2632.

Ciani, Emanuele, Louis Fréget, and Thomas Manfredi (2021). "Learning about Inequality and Demand for Redistribution: A Meta-Analysis of In-Survey Informational Experiments". *OECD Papers on Well-being and Inequalities* 02.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2007). "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects". *Economics of Education Review* 26(6), 673–682.

Corak, Miles (2013). "Income Inequality, Equality of Opportunity, and Intergenerational Mobility". *Journal of Economic Perspectives* 27(3), 79–102.

Cortes, Guido Matias (2016). "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data". *Journal of Labor Economics* 34(1), 63–105.

Coté, Stéphane, Julian House, and Robb Willer (2015). "High Economic Inequality Leads Higher-Income Individuals to be Less Generous". *Proceedings of the National Academy of Sciences of the United States of America* 112(52), 15838–15843.

Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz (2013). "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment". *Journal of Public Economics* 98, 100–112.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006). "Interpreting the Evidence on Life Cycle Skill Formation". *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek and Finis Welch. Vol. 1. Elsevier. Chap. 12, 697–812.

Currie, Janet, Henrik Kleven, and Esmée Zwiers (2020). "Technology and Big Data Are Changing Economics: Mining Text to Track Methods". *AEA Papers and Proceedings* 110, 42–48.

Cutler, David and Adriana Lleras-Muney (2006). "Education and Health: Evaluating Theories and Evidence". *NBER Working Paper 12352*.

Dachs, Bernhard (2018). *The Impact of New Technologies on the Labour Market and the Social Economy*. Brussels: EPRS.

Dahmann, Sarah C. (2017). "How Does Education Improve Cognitive Skills? Instructional Time Versus Timing of Instruction". *Labour Economics* 47, 35–47.

Dauth, Wolfgang, Sebastian Findeisen, Jens Suedekum, and Nicole Woessner (2021). "The Adjustment of Labor Markets to Robots". *Journal of the European Economic Association* 19(6), 3104–3153.

Davis, Jonathan M.V. and Sara B. Heller (2017). "Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs". *American Economic Review* 107(5), 546–550.

Dekker, Gerben W., Mykola Pechenizkiy, and Jan M. Vleeshouwers (2009). "Predicting Students Drop Out: A Case Study". *Proceedings of the 2nd International Conference on Educational Data Mining, EDM 2009, July 1-3, 2009. Cordoba, Spain*.

Dengler, Katharina and Britta Matthes (2018a). “Substituierbarkeitspotenziale von Berufen: Wenige Berufsbilder halten mit der Digitalisierung Schritt”. *IAB Kurzbericht 4/2018*.

– (2018b). “The Impacts of Digital Transformation on the Labour Market: Substitution Potentials of Occupations in Germany”. *Technological Forecasting and Social Change* 127, 304–316.

DeWit, David J., David DuBois, Gizem Erdem, Simon Larose, Ellen L. Lipman, and Renée Spencer (2016). “Mentoring Relationship Closures in Big Brothers Big Sisters Community Mentoring Programs: Patterns and Associated Risk Factors”. *American Journal of Community Psychology* 57(1-2), 60–72.

Dodin, Majed, Sebastian Findeisen, Lukas Henkel, Dominik Sachs, and Paul Schüle (2021). “Social Mobility in Germany”. *CESifo Working Paper 9200*.

Duquette, Nicolas J. and Enda P. Hargaden (2021). “Inequality and Giving”. *Journal of Economic Behavior & Organization* 186, 189–200.

Durante, Ruben, Louis Putterman, and Joël van der Weele (2014). “Preferences for Redistribution and Perception of Fairness”. *Journal of the European Economic Association* 12(4), 1059–1086.

Edlund, Jonas (1999). “Trust in Government and Welfare Regimes: Attitudes to Redistribution and Financial Cheating in the USA and Norway”. *European Journal of Political Research* 35, 341–370.

Eisenhauer, Philipp, James J. Heckman, and Edward Vytlacil (2015). “The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs”. *Journal of Political Economy* 123(2), 413–443.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). “Global Evidence on Economic Preferences”. *The Quarterly Journal of Economics* 133(4), 1645–1692.

Farbmacher, Helmut, Heinrich Kögel, and Martin Spindler (2021). “Heterogeneous Effects of Poverty on Attention”. *Labour Economics* 71, 102028.

Farbman, David A. (2015). “The Case for Improving and Expanding Time in School: A Review of Key Research and Practice”. *National Center on Time & Learning*.

Fehr, Dietmar, Johanna Mollerstrom, and Ricardo Perez-Truglia (2022a). “Your Place in the World: Relative Income and Global Inequality”. *American Economic Journal: Economic Policy* 14(4), 232–68.

Fehr, Dietmar, Daniel Müller, and Marcel Preuss (2022b). “(In-)Equality of Opportunity, Fairness, and Distributional Preferences”. *CESifo Working Paper 10001*.

Figlio, David, Paola Giuliano, Umut Özek, and Paola Sapienza (2019). “Long-Term Orientation and Educational Performance”. *American Economic Journal: Economic Policy* 11(4), 272–309.

Figlio, David, Kristian L. Holden, and Umut Ozek (2018). “Do Students Benefit from Longer School Days? Regression Discontinuity Evidence from Florida’s Additional Hour of Literacy Instruction”. *Economics of Education Review* 67, 171–183.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv (2017). "Distributional Preferences and Political Behavior". *Journal of Public Economics* 155, 1–10.

Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits (2015). "The Distributional Preferences of an Elite". *Science* 349(6254).

Fisman, Raymond, Shachar Kariv, and Daniel Markovits (2007). "Individual Preferences for Giving". *American Economic Review* 97(5), 1858–1876.

Fong, Christina (2001). "Social Preferences, Self-Interest, and the Demand for Redistribution". *Journal of Public Economics* 82(2), 225–246.

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (2018). "Mikrozensus 2018, SUF, Version 1". *Available online: https://www.forschungsdatenzentrum.de/de/10-21242-12211-2018-00-00-3-1-1; last accessed 14/09/2023*.

Freundl, Vera, Elisabeth Grewenig, Franziska Kugler, Philipp Lergetporer, Ruth Schüler, Katharina Wedel, Katharina Werner, Olivia Wirth, and Ludger Woessmann (2023). "The ifo Education Survey 2014–2021: A New Dataset on Public Preferences for Education Policy in Germany". *Jahrbücher für Nationalökonomie und Statistik* forthcoming.

Frey, Carl Benedikt and Michael A. Osborne (2017). "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114, 254–280.

Friedman, Jerome H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine". *The Annals of Statistics* 29(5), 1189–1232.

Friedman, Jerome H., Trevor J. Hastie, and Robert J. Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". *Journal of Statistical Software* 33(1).

Gaines, Brian J., James H. Kuklinski, Paul J. Quirk, Buddy Peyton, and Jay Verkuilen (2007). "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq". *The Journal of Politics* 69(4), 957–974.

García, Salvador, Julián Luengo, and Francisco Herrera (2015). *Data Preprocessing in Data Mining*. Vol. 72. Cham: Springer International Publishing.

Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Kirk Walters, Mengli Song, Seth Brown, Steven Hurlburt, Pei Zhu, Susan Sepanik, Fred Doolittle, and Elizabeth Warner (2010). "Middle School Mathematics Professional Development Impact Study: Findings After the First Year of Implementation". *NCEE 2010-4009*.

Gärtner, Karla (2002). "Differentielle Sterblichkeit - Ergebnisse des Lebenserwartungssurveys". *Zeitschrift für Bevölkerungswissenschaft* 27(2), 185–211.

Georgieff, Alexandre and Anna Milanez (2021). "What Happened to Jobs at High Risk of Automation?" *OECD Social, Employment and Migration Working Papers* 255.

Glewwe, Paul and Michael Kremer (2006). "Schools, Teachers, and Education Outcomes in Developing Countries". *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek and Finis Welch. Vol. 2. Elsevier. Chap. 16, 945–1017.

Goldhaber, Dan D. and Emily Anthony (2007). “Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching”. *The Review of Economics and Statistics* 89(1), 134–150.

Goldhaber, Dan D. and Dominic J. Brewer (1997). “Why Don’t Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity”. *Journal of Human Resources* 32(3), 505–523.

– (2000). “Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement”. *Educational Evaluation and Policy Analysis* 22(2), 129–145.

Golin, Marta and Christopher Rauh (2022). “The Impact of Fear of Automation”. *Cambridge Working Papers In Economics 2269*.

Goodman, Joshua (2014). “Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time”. *NBER Working Paper 20221*.

Graetz, Georg and Guy Michaels (2018). “Robots at Work”. *The Review of Economics and Statistics* 100(5), 753–768.

Grewenig, Elisabeth (2021). “School Track Decisions and Teacher Recommendations: Evidence from German State Reforms”. *ifo Working Papers 353*.

Grewenig, Elisabeth, Philipp Lergetporer, Lisa Simon, Katharina Werner, and Ludger Woessmann (2023). “Can Internet Surveys Represent the Entire Population? A Practitioners’ Analysis”. *European Journal of Political Economy* 78, 102382.

Grossman, Jean B., Christian S. Chan, Sarah E. O. Schwartz, and Jean E. Rhodes (2012). “The Test of Time in School-Based Mentoring: The Role of Relationship Duration and Re-Matching on Academic Outcomes”. *American Journal of Community Psychology* 49(1-2), 43–54.

Grossman, Jean B. and Jean E. Rhodes (2002). “The Test of Time: Predictors and Effects of Duration in Youth Mentoring Relationships”. *American Journal of Community Psychology* 30(2), 199–219.

Gust, Sarah, Eric A. Hanushek, and Ludger Woessmann (2022). “Global Universal Basic Skills: Current Deficits and Implications for World Development”. *NBER Working Paper 30566*.

Haaland, Ingar and Christopher Roth (2023). “Beliefs about Racial Discrimination and Support for Pro-Black Policies”. *The Review of Economics and Statistics* 105(1), 40–53.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023). “Designing Information Provision Experiments”. *Journal of Economic Literature* 61(1), 3–40.

Han, Jiawei, Micheline Kamber, and Jian Pei (2012). “Classification: Basic Concepts”. *Data Mining: Concepts and Techniques*. Ed. by Jiawei Han, Micheline Kamber, and Jian Pei. Elsevier. Chap. 8, 327–391.

Hanushek, Eric A. (1971). “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data”. *American Economic Review* 61(2), 280–288.

– (1986). “The Economics of Schooling: Production and Efficiency in Public Schools”. *Journal of Economic Literature* 24, 1141–1177.

Hanushek, Eric A. (1997). “Assessing the Effects of School Resources on Student Performance: An Update”. *Educational Evaluation and Policy Analysis* 19(2), 141–164.

– (2006). “School Resources”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek and Finis Welch. Vol. 2. Elsevier. Chap. 14, 865–908.

– (2020). “Education Production Functions”. *The Economics of Education*. Ed. by Steve Bradley and Colin Green. Elsevier. Chap. 13, 161–170.

Hanushek, Eric A., Lavinia Kinne, Philipp Lergetporer, and Ludger Woessmann (2022). “Patience, Risk-Taking, and Human Capital Investment Across Countries”. *The Economic Journal* 132(646), 2290–2307.

Hanushek, Eric A., Stephen J. Machin, and Ludger Woessmann (2016). “Editors’ Introduction”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann. Vol. 5. Elsevier, xiii–xiv.

Hanushek, Eric A., Marc Piopiunik, and Simon Wiederhold (2019). “The Value of Smarter Teachers”. *Journal of Human Resources* 54(4), 857–899.

Hanushek, Eric A. and Steven G. Rivkin (2006). “Teacher Quality”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek and Finis Welch. Vol. 2. Elsevier. Chap. 18, 1051–1078.

– (2010). “Generalizations about Using Value-Added Measures of Teacher Quality”. *American Economic Review* 100(2), 267–271.

Hanushek, Eric A. and Ludger Woessmann (2006). “Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries”. *The Economic Journal* 116(510), C63–C76.

– (2011). “The Economics of International Differences in Educational Achievement”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann. Vol. 3. Elsevier, 89–200.

– (2012). “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation”. *Journal of Economic Growth* 17(4), 267–321.

– (2015). *The Knowledge Capital of Nations: Education and the Economics of Growth*. CESifo book series. MIT Press.

– (2021). “Education and Economic Growth”. *Oxford Research Encyclopedia of Economics and Finance*.

Harris, Douglas N. and Tim R. Sass (2011). “Teacher Training, Teacher Quality and Student Achievement”. *Journal of Public Economics* 95(7–8), 798–812.

Hastie, Trevor J., Robert J. Tibshirani, and Jerome H. Friedman (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edition, corrected at 11th printing. New York, NY: Springer.

Hastie, Trevor J., Robert J. Tibshirani, and Martin Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

Hausner, Karl Heinz, Doris Söhnlein, Brigitte Weber, and Enzo Weber (2015). “Bessere Chancen mit mehr Bildung”. *IAB Kurzbericht 11/2015*.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith (1999). “The Economics and Econometrics of Active Labor Market Programs”. *Handbook of Labor Economics*. Ed. by Orley C. Ashenfelter and David Card. Vol. 3A. Elsevier. Chap. 31, 1865–2097.

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2006). “Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek and Finis Welsh. Vol. 1. Elsevier. Chap. 7, 307–458.

Heller, Sara B., Benjamin Jakubowski, Zubin Jelveh, and Max Kapustin (2023). “Machine Learning Can Predict Shooting Victimization Well Enough to Help Prevent It”. *NBER Working Paper 30170*.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack (2017). “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago”. *The Quarterly Journal of Economics* 132(1), 1–54.

Herrera, C., J. B. Grossman, T. J. Kauh, A. F. Feldman, J. McMaken, and L. Z. Jucovy (2007). *Making a Difference in Schools: The Big Brothers Big Sisters School-Based Mentoring Impact Study*. Philadelphia: Public/Private Ventures.

Heß, Pascal, Simon Janssen, and Ute Leber (2019). “Digitalisierung und berufliche Weiterbildung: Beschäftigte, deren Tätigkeiten durch Technologien ersetzbar sind, bilden sich seltener weiter”. *IAB Kurzbericht 16/2019*.

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk (2019). “Soft Commitments, Reminders, and Academic Performance”. *American Economic Journal: Applied Economics* 11(2), 114–142.

Holmlund, Helena, Sandra McNally, and Martina Viarengo (2010). “Does Money Matter for Schools?” *Economics of Education Review* 29(6), 1154–1164.

Howitt, Peter and Philippe Aghion (1998). “Capital Accumulation and Innovation as Complementary Factors in Long-Run Growth”. *Journal of Economic Growth* 3(2), 111–130.

Hoy, Christopher and Franziska Mager (2021). “Why Are Relatively Poor People Not More Supportive of Redistribution? Evidence from a Randomized Survey Experiment across Ten Countries”. *American Economic Journal: Economic Policy* 13(4), 299–328.

Huebener, Mathias, Susanne Kuger, and Jan Marcus (2017). “Increased Instruction Hours and the Widening Gap in Student Performance”. *Labour Economics* 47, 15–34.

Hvidberg, Kristoffer B., Claus T. Kreiner, and Stefanie Stantcheva (2022). “Social Positions and Fairness Views on Inequality”. *NBER Working Paper 28099*.

Innocenti, Stefania and Marta Golin (2022). “Human Capital Investment and Perceived Automation Risks: Evidence from 16 Countries”. *Journal of Economic Behavior & Organization* 195, 27–41.

Jackson, C. Kirabo (2018). “Does School Spending Matter? The New Literature on an Old Question”. *NBER Working Paper 25368*.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico (2016). “The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms”. *The Quarterly Journal of Economics* 131(1), 157–218.

Jackson, C. Kirabo, Cora Wigger, and Heyu Xiong (2021). “Do School Spending Cuts Matter? Evidence from the Great Recession”. *American Economic Journal: Economic Policy* 13(2), 304–335.

Jacob, Brian A. (2007). “The Challenges of Staffing Urban Schools with Effective Teachers”. *The Future of Children* 17(1), 129–153.

Jacob, Brian A. and Lars Lefgren (2004). “The Impact of Teacher Training on Student Achievement”. *Journal of Human Resources* 39(1), 50–79.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer (2021). “Worker Beliefs About Outside Options”. *NBER Working Paper 29623*.

Jakiela, Pamela (2015). “How Fair Shares Compare: Experimental Evidence From Two Cultures”. *Journal of Economic Behavior & Organization* 118, 40–54.

James, Gareth, Daniela Witten, Trevor J. Hastie, and Robert J. Tibshirani (2015). *An Introduction to Statistical Learning: With Applications in R*. Corrected at 6th printing. New York et al.: Springer.

Jarrett, Olga S., Darlene M. Maxwell, Carrie Dickerson, Pamela Hoge, Gwen Davies, and Amy Yetley (1998). “Impact of Recess on Classroom Behavior: Group Effects and Individual Differences”. *The Journal of Educational Research* 92(2), 121–126.

Jeffrey, Karen (2021). “Automation and the Future of Work: How Rhetoric Shapes the Response in Policy Preferences”. *Journal of Economic Behavior & Organization* 192, 417–433.

Jerrim, John, Luis Alejandro Lopez-Agudo, Oscar D. Marcenaro-Gutierrez, and Nikki Shure (2017). “What Happens When Econometrics and Psychometrics Collide? An Example Using the PISA Data”. *Economics of Education Review* 61, 51–58.

Jones, Charles, ed. (2002). *Introduction to Economic Growth*. 2nd edition. New York, USA: W.W. Norton.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger (2008). “What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City”. *Economics of Education Review* 27(6), 615–631.

Karadja, Mounir, Johanna Mollerstrom, and David Seim (2017). “Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution”. *The Review of Economics and Statistics* 99(2), 201–212.

Karcher, Michael J. (2005). “The Effects of Developmental Mentoring and High School Mentors’ Attendance on Their Younger Mentees’ Self-Esteem, Social Skills, and Connectedness”. *Psychology in the Schools* 42(1), 65–77.

Katz, Lawrence F. and Kevin M. Murphy (1992). “Changes in Relative Wages, 1963-1987: Supply and Demand Factors”. *The Quarterly Journal of Economics* 107(1), 35–78.

Kemper, Lorenz, Gerrit Vorhoff, and Berthold U. Wigger (2020). “Predicting Student Dropout: A Machine Learning Approach”. *European Journal of Higher Education* 10(1), 28–47.

Keusch, Florian and Chan Zhang (2017). “A Review of Issues in Gamified Surveys”. *Social Science Computer Review* 35(2), 147–166.

Khanna, Kabir and Gaurav Sood (2018). “Motivated Responding in Studies of Factual Learning”. *Political Behavior* 40(1), 79–101.

Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos (2007). “Data Mining Techniques for the Detection of Fraudulent Financial Statements”. *Expert Systems with Applications* 32(4).

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018). “Human Decisions and Machine Predictions”. *The Quarterly Journal of Economics* 133(1), 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015). “Prediction Policy Problems”. *American Economic Review* 105(5), 491–495.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). “Experimental Analysis of Neighborhood Effects”. *Econometrica* 75(1), 83–119.

Kluegel, James R. and Eliot R. Smith (1986). *Beliefs about Inequality: Americans’ Views of What Is and What Ought to Be*. Routledge.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff (2015). “Value-Added Modeling: A Review”. *Economics of Education Review* 47, 180–195.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk (2020). “The Formation of Prosociality: Causal Evidence on the Role of Social Environment”. *Journal of Political Economy* 128(2), 434–467.

Kotsiantis, S. B., C. Pierrakeas, and P. E. Pintelas (2004). “Predicting Students’ Performance in Distance Learning Using Machine Learning Techniques”. *Applied Artificial Intelligence* 18(5), 411–426.

Kupersmidt, Janis B., Kathryn N. Stump, Rebecca L. Stelter, and Jean E. Rhodes (2017). “Predictors of Premature Match Closure in Youth Mentoring Relationships”. *American Journal of Community Psychology* 59(1-2), 25–35.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva (2015). “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments”. *American Economic Review* 105(4), 1478–1508.

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

– (2020). “Expanding School Resources and Increasing Time on Task: Effects on Students’ Academic and Noncognitive Outcomes”. *Journal of the European Economic Association* 18(1), 232–265.

Lee, Se Woong and Eunjung Alice Lee (2020). “Teacher Qualification Matters: The Association between Cumulative Teacher Qualification and Students’ Educational Attainment”. *International Journal of Educational Development* 77.

Lergetporer, Philipp, Katharina Werner, and Ludger Woessmann (2020). “Educational Inequality and Public Policy Preferences: Evidence from Representative Survey Experiments”. *Journal of Public Economics* 188, 104226.

– (2021). “Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Representative Evidence from Adults and Adolescents”. *Economica* 88(351), 624–670.

List, John A., Azeem M. Shaikh, and Yang Xu (2019). “Multiple Hypothesis Testing in Experimental Economics”. *Experimental Economics* 22(4), 773–793.

Lochner, Lance J. (2011). “Nonproduction Benefits of Education: Crime, Health, and Good Citizenship”. *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann. Vol. 4. Elsevier. Chap. 2, 183–282.

– (2020). “Education and Crime”. *The Economics of Education*. Ed. by Steve Bradley and Colin Green. Elsevier. Chap. 9, 109–117.

Lucas, Adrienne M., Patrick J. McEwan, Moses Ngware, and Moses Oketch (2014). “Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda”. *Journal of Policy Analysis and Management* 33(4), 950–976.

Lucas, Robert E. (1988). “On the Mechanics of Economic Development”. *Journal of Monetary Economics* 22(1), 3–42.

Luttmer, Enzo F. P. and Monica Singhal (2011). “Culture, Context, and the Taste for Redistribution”. *American Economic Journal: Economic Policy* 3(1), 157–179.

Mankiw, N. Gregory, David Romer, and David N. Weil (1992). “A Contribution to the Empirics of Economic Growth”. *The Quarterly Journal of Economics* 107(2), 407–437.

Manning, Matthew, Gabriel T. W. Wong, Christopher M. Fleming, and Susanne Garvis (2019). “Is Teacher Qualification Associated with the Quality of the Early Childhood Education and Care Environment? A Meta-Analytic Review”. *Review of Educational Research* 89(3), 370–415.

Manski, Charles F. and Francesca Molinari (2010). “Rounding Probabilistic Expectations in Surveys”. *Journal of Business & Economic Statistics* 28(2), 219–231.

Marcotte, Dave E. (2007). “Schooling and Test Scores: A Mother-Natural Experiment”. *Economics of Education Review* 26(5), 629–640.

Marečková, Jana and Winfried Pohlmeier (2019). “How Well Can Noncognitive Skills Predict Unemployment? A Machine Learning Approach”. *Mimeo*.

Martin, M. O., I.V.S. Mullis, and M. Hooper, eds. (2016). *Methods and Procedures in TIMSS 2015*. Chestnut Hill, MA, USA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Matthewes, Sönke Hendrik (2021). “Better Together? Heterogeneous Effects of Tracking on Student Achievement”. *The Economic Journal* 131(635), 1269–1307.

McCall, Leslie, Derek Burk, Marie Laperrière, and Jennifer A. Richeson (2017). “Exposure to Rising Inequality Shapes Americans’ Opportunity Beliefs and Policy Support”. *Proceedings of the National Academy of Sciences of the United States of America* 114(36), 9593–9598.

Meroni, Elena Claudia and Giovanni Abbiati (2016). “How Do Students React to Longer Instruction Time? Evidence from Italy”. *Education Economics* 24(6), 592–611.

Metzler, Johannes and Ludger Woessmann (2012). “The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation”. *Journal of Development Economics* 99(2), 486–496.

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

Molnar, Christoph (2021). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*.

Montenegro, Claudio E. and Harry Anthony Patrinos (2014). “Comparable Estimates of Returns to Schooling Around the World”. *Policy Research Working Paper 7020*.

Montobbio, Fabio, Jacopo Staccioli, Maria Virgillito, and Marco Vivarelli (2022). “The Empirics of Technology, Employment and Occupations: Lessons Learned and Challenges Ahead”. *IZA Discussion Paper 15731*.

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa (2021). “Job Seekers’ Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias”. *American Economic Review* 111(1), 324–363.

Mullainathan, Sendhil and Jann Spiess (2017). “Machine Learning: An Applied Econometric Approach”. *Journal of Economic Perspectives* 31(2), 87–106.

Müller, Normann and Felix Wenzelmann (2020). “Berufliche Weiterbildung – Teilnahme und Abstinenz”. *Zeitschrift für Weiterbildungsforschung* 43(1), 47–73.

Mummolo, Jonathan and Erik Peterson (2019). “Demand Effects in Survey Experiments: An Empirical Assessment”. *American Political Science Review* 113(2), 517–529.

National Center on Time & Learning (2017). “Why Time Matters”. *Available online: http://www.timeandlearning.org/?q=why-time-matters-0; last accessed: 12/09/2023*.

National Education Panel Study (NEPS) (2020). *NEPS-Startkohorte 4: Klasse 9 (SC4 11.0.0)*. Bamberg: Leibniz Institute for Educational Trajectories (LIfBi).

Nedelkoska, Ljubica and Glenda Quintini (2018). “Automation, Skills Use and Training”. *OECD Social, Employment and Migration Working Papers 202*.

Nickell, Stephen (1979). “Estimating the Probability of Leaving Unemployment”. *Econometrica* 47(5), 1249–1266.

– (2004). “Poverty and Worklessness in Britain”. *The Economic Journal* 114(494), C1–C25.

Nilsen, Trude, Ronny Scherer, and Sigrid Blömeke (2018). “The Relation of Science Teachers’ Quality and Instruction to Student Motivation and Achievement in the 4th and 8th Grade: A

Nordic Perspective". *Northern Lights on TIMSS and PISA 2018*. Ed. by Anita Wester. Nordisk Ministerråd. Chap. 3, 61–94.

Norton, Michael I. and Dan Ariely (2011). "Building a Better America - One Wealth Quintile at a Time". *Perspectives on Psychological Science* 6(1), 9–12.

OECD (1998). *Technology, Productivity and Job Creation: Best Policy Practices*. The OECD Jobs Strategy. Paris: OECD Publishing.

– (2014). "Education Indicators in Focus: How Much Time Do Primary and Lower Secondary Students Spend in the Classroom?" *Education Indicators in Focus* 22.

– (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.

– (2018). *Equity in Education*. Paris: OECD Publishing.

– (2020). *Education Policy Outlook: Germany*. Paris: OECD Publishing.

– (2021). *Continuing Education and Training in Germany: Getting Skills Right*. Paris: OECD Publishing.

Oreopoulos, Philip, Robert S. Brown, and Adam M. Lavecchia (2017). "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students". *Journal of Political Economy* 125(4), 947–984.

Osberg, Lars and Timothy Smeeding (2006). ""Fair" Inequality? Attitudes Toward Pay Differentials: The United States in Comparative Perspective". *American Sociological Review* 71(3), 450–473.

Osiander, Christopher and Gesine Stephan (2020). "Was beeinflusst die Weiterbildungsbereitschaft von Beschäftigten? Befunde aus einer Vignettenstudie". *Industrielle Beziehungen* 27(3), 336–359.

Papay, John P. and Matthew A. Kraft (2015). "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement". *Journal of Public Economics* 130, 105–119.

Patall, Erika A., Harris Cooper, and Ashley Batts Allen (2010). "Extending the School Day or School Year: A Systematic Review of Research (1985-2009)". *Review of Educational Research* 80(3), 401–436.

Paulus, Wiebke and Britta Matthes (2013). "The German Classification of Occupations 2010 – Structure, Coding and Conversion Table". *FDZ-Methodenbericht 08/2013*.

Payne, A. Abigail and Justin Smith (2015). "Does Income Inequality Increase Charitable Giving?" *The Canadian Journal of Economics* 48(2), 793–818.

Piketty, Thomas (1995). "Social Mobility and Redistributive Politics". *The Quarterly Journal of Economics* 110(3), 551–584.

Podgursky, Michael J. and Matthew G. Springer (2007). "Teacher Performance Pay: A Review". *Journal of Policy Analysis and Management* 26(4), 909–950.

Puleston, Jon (2011). "Improving Online Surveys". *International Journal of Market Research* 53(4), 557–562.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth (2018). "Measuring and Bounding Experimenter Demand". *American Economic Review* 108(11), 3266–3302.

Raposa, Elizabeth B., Adar Ben-Eliyahu, Lauren E. W. Olsho, and Jean E. Rhodes (2019). "Birds of a Feather: Is Matching Based on Shared Interests and Characteristics Associated with Longer Youth Mentoring Relationships?" *Journal of Community Psychology* 47(2), 385–397.

Resnjanskij, Sven, Jens Ruhose, Simon Wiederhold, Ludger Woessmann, and Katharina Wedel (2024). "Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects". *Journal of Political Economy* forthcoming.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). "Teachers, Schools, and Academic Achievement". *Econometrica* 73(2), 417–458.

Rivkin, Steven G. and Jeffrey C. Schiman (2015). "Instruction Time, Classroom Quality, and Academic Achievement". *The Economic Journal* 125(588), F425–F448.

Rockoff, Jonah E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data". *American Economic Review* 94(2), 247–252.

Rockoff, Jonah E., Brian Jacob, Thomas Kane, and Douglas Staiger (2011). "Can You Recognize an Effective Teacher When You Recruit One?" *NBER Working Paper 14485*.

Rodríguez-Planas, Núria (2012). "Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States". *American Economic Journal: Applied Economics* 4(4), 121–139.

Roemer, John E. (2004). "Equal Opportunity and Intergenerational Mobility: Going Beyond Intergenerational Income Transition Matrices". *Generational Income Mobility in North America and Europe*. Ed. by Miles Corak. Cambridge University Press. Chap. 3, 48–57.

Romano, Joseph P. and Michael Wolf (2010). "Balanced Control of Generalized Error Rates". *The Annals of Statistics* 38(1), 598–633.

Romer, Paul M. (1990). "Endogenous Technological Change". *Journal of Political Economy* 98(5), 71–102.

Roth, Christopher and Johannes Wohlfart (2018). "Experienced Inequality and Preferences for Redistribution". *Journal of Public Economics* 167, 251–262.

Ruan, Jianfei, Zheng Yan, Bo Dong, Qinghua Zheng, and Buyue Qian (2019). "Identifying Suspicious Groups of Affiliated-Transaction-Based Tax Evasion in Big Data". *Information Sciences* 477.

Sandner, Malte (2015). "The Effects of High-Quality Student Mentoring". *Economics Letters* 136, 227–232.

Sansone, Dario (2019). "Beyond Early Warning Indicators: High School Dropout and Machine Learning". *Oxford Bulletin of Economics and Statistics* 81(2), 456–485.

Sass, Tim R. (2015). "Licensure and Worker Quality: A Comparison of Alternative Routes to Teaching". *Andrew Young School of Policy Studies Research Paper Series 13-09*.

Schmidpeter, Bernhard and Rudolf Winter-Ebmer (2021). "Automation, Unemployment, and the Role of Labor Market Training". *European Economic Review* 137(103808).

Schmillen, Achim and Heiko Stüber (2014). "Bildung lohnt sich ein Leben lang". *IAB Kurzbericht 1/2014*.

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

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Woessmann (2008). "Education Policy and Equality of Opportunity". *Kyklos* 61(2), 279–308.

Schwerdt, Guido and Amelie C. Wuppermann (2011). "Is Traditional Teaching Really all that Bad? A Within-Student Between-Subject Approach". *Economics of Education Review* 30(2), 365–379.

Seyda, Susanne and Beate Placke (2020). *IW-Weiterbildungserhebung 2020: Weiterbildung auf Wachstumskurs*. Köln: Institut der deutschen Wirtschaft Köln e.V.

Shuls, James V. and Julie R. Trivitt (2015). "Teacher Qualifications and Productivity in Secondary Schools". *Journal of School Choice* 9(1), 49–70.

Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth". *The Quarterly Journal of Economics* 70(1), 65.

Spiewak, Martin (2018). "Es wäre gerechter, die Grundschulzeit zu verlängern". *Die Zeit. Available online: https://www.zeit.de/gesellschaft/2018-05/chancengleichheit-herkunft-elternhaus-universitaet-akademikerfamilie, last accessed: 25/01/2022*.

Staiger, Douglas O. and Jonah E. Rockoff (2010). "Searching for Effective Teachers with Imperfect Information". *Journal of Economic Perspectives* 24(3), 97–118.

Stanat, Petra, Stefan Schipolowski, Camilla Rjosk, Sebastian Weirich, and Nicole Haag, eds. (2017). *IQB-Bildungstrend 2016: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster and New York: Waxmann.

Stantcheva, Stefanie (2022). "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible". *NBER Working Paper 30527*.

Statistik der Bundesagentur für Arbeit (2023a). "Berufe auf einen Blick". *Available online: https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Berufe-auf-einen-Blick/Berufe-auf-einen-Blick-Anwendung-Nav.html, last accessed: 12/09/2023*. Ed. by Statistik der Bundesagentur für Arbeit.

– (2023b). "Datenbank Beschäftigung". *Available online: https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Datenbanken/Datenbanken-BST-Nav.html, last accessed: 12/09/2023*. Ed. by Statistik der Bundesagentur für Arbeit.

Statistisches Bundesamt (2018). *Schulen auf einen Blick – Ausgabe 2018*. Wiesbaden: Statistisches Bundesamt (Destatis).

Swan, Trevor W. (1956). "Economic Growth and Capital Accumulation". *Economic Record* 32(2), 334–361.

TIMSS (2017). *Supplement 1: International Version of the TIMSS 2015 Context Questionnaires*. International Association for the Evaluation of Educational Achievement.

– (2019). "About TIMSS & PIRLS International Study Center". *Available online: https://timss.bc.edu/about.html, last accessed: 12/09/2023*.

UNESCO (2023). "Instructional Time Definition". *Available online: http://www.ibe.unesco.org/en/glossary-curriculum-terminology/i/instructional-time, last accessed: 12/09/2023*.

United Nations (2014). *World Economic Situation and Prospects 2014 - Country Classification*.

van den Berg, Gerard J., Christine Dauth, Pia Homrighausen, and Gesine Stephan (2019). "Informing Employees in Small and Medium Sized Firms About Training: Results of a Randomized Field Experiment". *IAB-Discussion Paper 22-2019*.

Varian, Hal R. (2014). "Big Data: New Tricks for Econometrics". *The Journal of Economic Perspectives* 28(2), 3–28.

Wager, Stefan and Susan Athey (2018). "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests". *Journal of the American Statistical Association* 113(523), 1228–1242.

Werner, Katharina, Vera Freundl, Franziska Kugler, Philipp Lergetporer, Katharina Wedel, and Ludger Woessmann (2022). "Deutsche befürworten Weiterbildung, um mit dem Strukturwandel Schritt zu halten: Ergebnisse des ifo Bildungsbarometers 2022". *ifo Schnelldienst* 75(9), 56–69.

Wiswall, Matthew (2013). "The Dynamics of Teacher Quality". *Journal of Public Economics* 100, 61–78.

Witten, I. H., Christopher J. Pal, Eibe Frank, and Mark A. Hall (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. 4th. Cambridge, MA: Morgan Kaufmann.

Woessmann, Ludger (2003). "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence". *Oxford Bulletin of Economics and Statistics* 65(2), 117–170.

– (2006). "International Evidence on Expenditure and Class Size: A Review". *Brookings Papers on Education Policy* 2006(1), 245–272.

Woessmann, Ludger, Florian Schoner, Vera Freundl, and Franziska Pfaehler (2023). "Der ifo -"Ein Herz für Kinder" - Chancenmonitor: Wie (un-)gerecht sind die Bildungschancen von Kindern aus verschiedenen Familien in Deutschland verteilt?" *ifo Schnelldienst* 76(4), 29–47.

Woessmann, Ludger and Martin West (2006). "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS". *European Economic Review* 50(3), 695–736.

Wu, Derek (2020). "Disentangling the Effects of the School Year from the School Day: Evidence from the TIMSS Assessments". *Education Finance and Policy* 15(1), 104–135.

Wyness, Gill, Lindsey Macmillan, Jake Anders, and Catherine Dilnot (2023). "Grade Expectations: How Well Can Past Performance Predict Future Grades?" *Education Economics* 31(4), 397–418.

Zhang, Baobao (2022). "No Rage Against the Machines". *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. Ed. by Vincent Conitzer, John Tasioulas, Matthias Scheutz, Ryan Calo, Martina Mara, and Annette Zimmermann. New York, NY, USA: ACM, 856–866.

Zhou, Victor (2019). "A Simple Explanation of Gini Impurity". *Available online: https://victorzhou.com/blog/gini-impurity/, last accessed: 12/09/2023*.

Zou, Hui (2006). "The Adaptive Lasso and Its Oracle Properties". *Journal of the American Statistical Association* 101(476), 1418–1429.