INAUGURAL DISSERTATION 2016

Students' Performance and Behavior in Elementary and Higher Education

THREE ESSAYS
IN THE ECONOMICS
OF EDUCATION

Stefanie P. Herber
University of Bamberg

BAMBERG GRADUATE SCHOOL OF SOCIAL SCIENCES

Students' Performance and Behavior in Elementary and Higher Education

Three Essays in the Economics of Education

Inaugural-Dissertation
zur Erlangung der Doktorwürde
— Doctor rerum politicarum —
an der Fakultät Sozial- und Wirtschaftswissenschaften
der Otto-Friedrich-Universität Bamberg

vorgelegt von

Stefanie P. Herber, M. Sc.

2016

Kumulative Dissertation

Erstgutachter:	Prof. Dr. Guido Heineck
Zweitgutachterin:	Prof. Dr. Silke Anger
Drittgutachter:	Prof. Dr. Guido Schwerdt

Tag der mündlichen Prüfung:	08.08.2016
Dissertationsort:	Bamberg

Acknowledgments

I have enjoyed reading acknowledgments in dissertations for years because they are often probably the most beautiful and most pathetic pieces devoted to a public most economists—ill-reputed as overly rational—are ever going to write. Writing my own acknowledgments feels like putting a temporary period to the end of a cycle of ups and downs. During this time, I have received enormous support from many people who already were, or have become, an important part of my life.

First of all, I would like to thank my supervisor Guido Heineck for his ongoing support, encouragement, and confidence in my work. I appreciated his valuable scientific guidance, but also his personal assistance and support that carried me through the past three years. I am also indebted to my first co-supervisor, Silke Anger, for helpful comments and advice on my work. Moreover, I want to thank Guido Schwerdt for kindly agreeing to be the second co-supervisor. In addition, I am indebted to my co-authors Michael Kalinowski and Johanna Sophie Quis for their collaboration on the second and third chapter of this dissertation.

I would also like to thank the Bamberg Graduate School of Social Sciences (BAGSS)—not only for accepting and financing me as a doctoral student, but also for the great dedication of Miriam Schneider, Marc Scheibner, and Katrin Bernsdorff. Furthermore, I owe gratitude to Steffen Schindler, Corinna Kleinert, and all other participants in the BAGSS and the economics colloquium for offering me a platform to bring my work up for discussion and taking the time to comment on my papers.

Many of my colleagues at BAGSS have also become friends who made bad days more bearable and good days even more cheerful. Marion Fischer-Neumann, thank you for sharing an office with me and always having a friendly ear. Martin, I am happy we could always share coffee, beer, and sorrows. Johanna Sophie, thank you for being the best travel companion ever and providing me with a more relaxed view of things. And, last but definitely not least, thank you, Irina Hondralis and Gundula Zoch, for consulting me in dissertation-related questions, for our restorative lunches and dinners, and for always being by my side.

I am also indebted to several other people from in- and outside Bamberg with whom I could discuss several aspects of my work, especially Julia Graf, Zoltán J. Juhász, Daniel Kamhöfer, Frauke Peter, Klaus Pforr, Tobias Rausch, and Ulrich Schroeders. Moreover, I am deeply grateful to count on close friends who were always there, especially Robert Biene and Lars Uekötter, Jutta Herrmann, and Carolin Vöckler.

Apart from supervisors, colleagues, and friends, I want to thank my parents for always believing in me and having paved the way for my graduate career—by now, I understand the returns to your early human capital investments in me even better.

Finally, I am deeply grateful to Kevin. I would not have arrived at this point without your unconditional support, continuous motivation, and loving encouragement.

Thank you all!

Bamberg, March 31, 2016
Stefanie P. Herber

Contents

List of Figures

List of Tables

List of abbreviations

AME	Average marginal effects
BA	Bachelor’s degree
BAföG	Federal Education and Training Assistance Act
BFW	Begabtenförderungswerke
DST	Daylight saving time
FAFSA	Free Application for Federal Student Aid
GDR	German Democratic Republic
GPA	Grade point average
HLM	Hierarchical linear model
IEA	International Association for the Evaluation of Educational Achievement
ITT	Intent-to-treat
IV	Instrumental variable
MA	Master’s degree
NTU	Non-take-up rate
OLS	Ordinary least squares
PIRLS	Progress in International Reading Literacy Study
SAT	Scholastic Aptitude Test

SOEP	Socio-Economic Panel
ST	Standard time
STEM	Science, technology, engineering, and mathematics
TIMSS	Trends in International Mathematics and Science Study
TSLS	Two-stage least-squares

Chapter 1

Introduction

The first studies in the economics of education were concerned with recognizing human capital as an investment good (Schultz, 1959, 1961; Becker, 1993) and to overcome the fact that "our knowledge about national wealth is almost wholly restricted to the non-human components, that is, to reproducible physical capital and land" (Schultz, 1959, p. 110). Nowadays, investments in education are seen as "one of the top priority policy areas of governments around the world [...] [and] an essential element in global economic competition" (Hanushek, 2009, p. 39). Accordingly, there is large interest by both policy makers and researchers with respect to the determinants of educational achievements, the outcomes of education, and the lessons to be learned from and for educational policies.

Generally, educational policy aims at both remedying market imperfections and establishing equality of opportunity. Market imperfections with respect to education are mainly produced by the fact that education exerts positive externalities, resulting in individual human capital investments below the social optimum. Moreover, imperfect information about costs and benefits of education provides another rationale for intervention—both for efficiency and equity reasons. Apart from that, equity considerations themselves give rise to public involvement in the provision and financing of education. Therefore, the common ground is typically to secure equality of opportunity, so that all people have equal opportunities to access education—irrespective of their socio-economic background.

This dissertation consists of two separate parts. The first part seizes on the determinants of educational achievement, whereas the second part is concerned with students' behavior towards federal interventions aiming at achieving efficiency and equality of opportunity.

The first part of this dissertation builds on the main idea that economists model the determinants of educational achievement in an education production function where student and family characteristics, but also institutional and school factors, influence test scores. Institutional factors comprise the form of funding, the autonomy of schools, and age of tracking, while school-related factors include class size, instruction time, and teacher quality.

Among the school-related factors—and probably even among all topics ever studied in the economics of education—, class size has received most attention by scholars and policy makers. Research of the past decades has shown that class size reductions are a very costly, though relatively ineffective, means to improve students' performance (see Hanushek (1999) or Hanushek (2006) for a review of the literature).

A school-level factor that has received far less scientific attention is school starting time. Though parents, teachers, and the media favor later school starting times with well rested students, the question whether ringing the school bell later increases students' school performance is still unresolved. On the one hand, circadian rhythms are tied to the light-dark cycle and getting up early in the dark might be related to shorter bed hours. Many studies suggest, indeed, a positive correlation between hours slept and grades in middle and high school (Wolfson and Carskadon, 2003; Shochat et al., 2014).

Yet, on the other hand, recent evidence on whether later school start times cause increased achievements yields mixed results (Edwards, 2012; Carrell et al., 2011; Hinrichs, 2011; Heissel and Norris, 2015).

Johanna Sophie Quis, Guido Heineck, and I take a broader perspective to investigate the relationship between sleep and students' performance in **Chapter 2** of this dissertation: We provide first evidence on whether the spring transition from standard to daylight saving time (DST) induces short-run consequences on elementary school students' test performance. Based on the public discussion and the pertinent literature from medicine, psychology, and biology, we hypothesize that elementary school children might suffer from sleep deprivation in the week after the clocks were advanced by one hour.

We exploit the fact that six European states collected data on more than 22,000 students in the *Trends in International Mathematics and Science Study* (TIMSS) and the *Progress in International Reading Literacy Study* (PIRLS) during the transition to DST in spring 2011. In a regression discontinuity design, we compare the performance of students randomly allocated to testing dates in the week before the clock advance with those allocated to testing dates in the week after the clock advance.

Our estimates for the DST-effect are very small in magnitude and not statistically significantly different from zero. Therefore, our results challenge the prevailing public opinion that daylight saving time should be abandoned because of its detrimental effects on school children's performance.

The second part of this dissertation shifts the focus to equality of opportunity in higher education and students' reaction to offers of different forms of student financial aid.

Student financial aid aims at overcoming credit constraints resulting from capital market imperfections: Human capital theory (Becker, 1993) argues that human capital cannot serve as a collateral for a loan as "Free men are not for sale" (Schultz, 1959, p. 111) and lenders cannot force graduates to deploy their full potential productively to prevent moral hazard. Therefore, access to higher education will depend on parents' assets if students are credit constrained and student financial aid is unavailable. Although most industrialized countries provide student financial aid to secure equality of opportunity, the intergenerational persistence in educational attainment is still high (Hertz et al., 2008; Heineck and Riphahn, 2009; Riphahn and Schieferdecker, 2010; Blossfeld et al., 2016). The results of Cameron and Heckman (1998, 2001) and Carneiro and Heckman (2002) provide a potential explanation for this finding: The authors argue that only long-run factors such as the parents' permanent income and endowment restrict the transition to higher education, and that short-run liquidity constraints are no relevant obstacle. According to them, intergenerational persistence of educational attainment is an artifact of a failure to account for the high unobserved ability or motivation of those students who make their way successfully through the educational system. Once these long-run constraints and dynamic sample selection are accounted for, they find no effect of parental income on US students' access to higher education in the *National Longitudinal Survey of Youth*. They argue that mitigating short-run liquidity constraints cannot increase enrollment and graduation rates substantially.

The findings of Cameron and Heckman (1998, 2001) are, however, challenged by several more recent studies: Belley and Lochner (2007), for example, use the same database but draw upon cohorts born 20 years later. While they can replicate Heckman and co-authors' findings for the older cohorts, they find a large increase in the impact of family income on college enrollment for the younger cohort, even after long-run factors have been accounted for. Furthermore, the impact of parents' income on completion of at least 2 years of college persists and is equally large. The authors do not account for sample selection, though.

The evidence for Germany is limited. Nevertheless, Riphahn and Schieferdecker (2010) investigate the importance of parental background factors in the transition from high school to university in Germany. With data from the *German Socio-Economic Panel Study* (SOEP), they show that parental income remains a large and significant predictor of enrollment probabilities, even though long-run characteristics such as parental education and sample selection into the pool of high school graduates are accounted for.

With respect to future governmental intervention, it is important to understand why the inter-generational educational mobility to German universities is relatively low (OECD, 2014, p. 93), although higher education institutions do not charge tuition, and student aid transfers based on the Federal Education and Training Assistance Act (BAföG) provide lucrative funding that should facilitate studying for students from low-income families.

Our analysis in **chapter 3** is an attempt to offer new insights into this matter. Michael Kalinowski and I study whether low-income students eligible to receive BAföG do indeed claim their benefits or whether features in the design of the federal aid scheme prevent students from claiming their need-based student aid amounts.

We construct a microsimulation model for the SOEP 2002–2013 to estimate the respective aid amounts students would have received, had they filed an application for need-based aid. The results indicate that about two fifths of the eligible low-income students do not take up their entitlements.

In a second step, we consider several potential explanatory factors shedding light on the reasons for students' non-take-up of BAföG. More specifically, we employ instrumental variable techniques and a sample selection model to investigate whether different utilities from claiming, information constraints, parents' claiming of other welfare benefits, cultural differences, and time-inconsistent preferences can explain students' behavior. We find that the expected duration of benefit receipt and students' financial need are inversely related to non-take-up. Nevertheless, increasing benefits by 10% decreases non-take-up by only 4.1% on average. In addition, students who can draw upon past experience of older siblings in filing the complex applications for BAföG are considerably more likely to claim the benefits. Therefore, these findings provide evidence that information asymmetries contribute to explaining why student financial aid may not work as intended. Moreover, we include a variable indicating whether parents lived in the former socialist East Germany before reunification to proxy inherited preferences about the welfare state. We find that significantly more students socialized in East Germany choose to take up student aid compared to similar West German students. Lastly, our results suggest that, although BAföG as a combination of grant and zero-interest loan is by construction rather profitable, it seems to favor debt aversion of highly impulsive and impatient students in line with the predictions from the "Economic Theory of Self-Control" (Thaler and Shefrin, 1981).

Although we abstain from claiming causality for the mechanisms we investigate, our results provide indications that the German need-based student financial aid scheme may not be suited to provide equality of opportunity if non-claiming students have to spend a considerable time on working to earn their living. Students who have to work many hours and can devote less resources to studying are, consequently, more susceptible to prolong studying (Avdic and Gartell, 2015), perform worse academically (Stinebrickner and Stinebrickner, 2003; Callender, 2008), or drop out without a degree (Triventi, 2014). The students we are analyzing have already made their way to university. Nevertheless, dropping out is harmful to equality because of the non-trivial monetary returns accruing from completing higher education, the "sheepskin effect" (Heckman et al., 2006, e.g.). It remains an open question to which extent the factors fostering higher education students' decisions not to claim also carry over to high school graduates. It is, however, plausible to assume that high school students' level of information about BAföG and their ability to cope with the complex paperwork is, if anything, lower than that of higher education students. The high complexity of BAföG and its design might, therefore, also dissuade a significant share of credit-constrained students from studying at all. Bettinger et al. (2012) substantiate this assumption for the US. They show that assisting low-income families in filling out student aid application forms increases aid take-up, college enrollment, and completion substantially.

While chapter 3 studies federal need-based aid, the subsequent chapter 4 is devoted to the analysis of federal merit-based aid. The main goal of merit-based scholarships is not to provide equal access for all students but rather to efficiently promote the most talented ones. This efficiency goal, currently budgeted with almost EUR 244 million of fiscal revenue in Germany, implies that the most talented students, irrespective of their social background, should receive a scholarship.

Yet, two thirds of all German merit-based aid holders come from families where at least one parent achieved a college degree, though students of academic background only make up half of the overall student population (Middendorff et al., 2009, p. 24). The magnitude of this difference is surprising, considering that differences in university grades between students from college-experienced and college-inexperienced families are very small (Delaney et al., 2011; Aspelmeier et al., 2012). Furthermore, the findings that information asymmetries even plague students' claiming of broad and therefore well-known need-based aid (see chapter 3 but also Dynarski and Scott-Clayton (2006); King (2006); Bettinger et al. (2012)) make it very likely that information asymmetries also play a role with respect to merit-based aid: Only 1% of all German higher education students receive a scholarship so that the mere existence of this funding possibility might be unknown to many students from families unacquainted with higher education. Moreover, while eligibility for need-based aid is unequivocally regulated by law, the eligibility criteria to successfully apply for merit-based aid are vague and vary between the privately-owned foundations distributing the federal funds to students. Insider information and encouragement from parents might thus be all the more helpful and necessary for students to consider applying. If information asymmetries kept eligible students off applying, scholarships would not meet the efficiency goal to further promote the talent of promising students.

Although only 1% of all German students are funded by merit-based scholarships, investigating social selectivity in the German merit-based aid system is important: Besides the risk of talent loss, being funded by a scholarship in Germany opens up access to several other non-monetary privileges such as mentoring and support by influential previous scholarship holders, and serves as a strong signal in the curriculum vitae of those who succeeded in the highly competitive selection procedure. A social selective scholarship system based on a considerable amount of public funds raises thus also normative equality concerns.

Chapter 4 explores whether information asymmetries between students able to draw upon the guidance of at least one college-experienced parent ("academic students") and those from families where no one studied ("non-academic students") are one of the causes for different application probabilities.

There are two reasons for why I focus on students' applications for scholarships rather than on their successes. First, I want to assess which share of the non-academic students' under-representation in the scholarship body is caused by their scant or distorted information about the scholarship system. I cannot isolate this channel when investigating success probabilities because many more factors influence the scholarship award. On the one hand, for example, scholarship providers may (consciously or unconsciously) discriminate against "educationally deprived" students. On the other hand, these students' performance in the assessment centers may be under threat if they feel stereotyped as a minority (Steele et al., 2002). Second, even if scholarship foundations are able to objectively select the best students, their choice is limited to the pool of applicants. Therefore, from a policy perspective, increasing the share of eligible applicants from non-academic homes provides the basis to secure an efficient allocation of funds.

A causal effect of information asymmetries could not be isolated even if data on students deciding for or against applying for a scholarship were available. Therefore, I conducted a web-based field experiment between the years 2013 and 2015, based on a sample of more than 5,000 German higher education students. I randomly assigned students either to the control group or one of two treatment groups. The first treatment group read a general primer about scholarships, based on the scholarship information publicly available. In the second treatment group, participants additionally read an interview with a real, current scholarship holder. He or she provided tailored information on the detailed application process and probabilities of success. To ease identification, the scholarship holder resembled the participant in several characteristics, acting as a role model.

At baseline, the results indicate indeed that non-academic students apply less often than students from academic homes, keeping a range of eligibility characteristics constant. Of the non-applicants at baseline, non-academic students are moreover significantly worse informed, suggesting that the decision to abstain from applying might not be well-grounded. Accordingly, both treatments increase the knowledge of non-academic students about scholarships significantly. Yet, only the role model treatment increases non-academic students' application probabilities for federally funded merit-based scholarships significantly. Information asymmetries are therefore one factor why non-academic students are underrepresented in the German scholarship system. Moreover, the public information about scholarships currently provided online is not sufficient to compensate for the existent knowledge differentials by parental background. Interpreting the role model treatment as a simulation of the custom-fit insider information many academic students can access through their parents, the decisive information is a glance behind the scenes and the assurance that a similar person made it. As a consequence, establishing mentoring programs at schools is a promising and inexpensive endeavor to increase both efficiency and equality of opportunity in the merit aid system.

Chapter 2

Does the Transition into Daylight Saving Time Affect Students' Performance?

Stefanie P. Herber, Johanna Sophie Quis, and Guido Heineck

2.1 Introduction

80 countries around the world[1] are currently exposed to a shift in sleep patterns twice a year when they switch between daylight saving time (DST) and standard time (ST): In the northern hemisphere, clocks are set forward by one hour in spring to DST and set backward by one hour in fall to ST. While the phase delay in fall rewards us with an additional hour of sleep, the phase advance in spring implies that we have to get up one hour earlier—while the sunlight lags one hour behind.

Ever since its first introduction, the change to DST has been critically discussed. Germany and Austria-Hungary introduced DST in 1916[2] in order to save energy and to better match sleep-wake cycles with daylight times. Various recent studies challenge that DST saves energy (e.g., Kellogg and Wolff, 2008; Aries and Newsham, 2008; Kotchen and Grant, 2011; Sexton and Beatty, 2014). Another strand of research discusses whether the shift to DST increases traffic and work-related accidents (e.g., Hicks et al., 1983; Barnes and Wagner, 2009) or not (e.g., Ferguson et al., 1995; Lahti et al., 2011), influences stock market returns (e.g., Kamstra et al., 2000) or not (e.g., Gregory-Allen

[1] Data compiled from the CIA World Factbook Central Intelligence Agency (2013).

[2] Reichsgesetzblatt (RGBI) 1916. Bekanntmachung über die Vorverlegung der Stunden während der Zeit vom 1. Mai bis 30. September 1916, RGBl 1916, p. 243.

et al., 2010), affects individuals' subjective well-being negatively (Kountouris and Remoundou, 2014; Kuehnle and Wunder, 2015), or might even increase the risk of heart attacks (e.g., Janszky et al., 2012) or not (e.g., Sandhu et al., 2014), to name just some.

So far, no clear conclusions can be drawn as to whether DST is indeed harmful enough to affect outcomes measurably and whether its potential costs outweigh its supposed benefits. Most previous studies suffer from small sample sizes (as already noted by Gregory-Allen et al., 2010) or fail to control for unobserved structural differences before and after the time change or between DST- and non-DST-countries. Nevertheless, a recent representative survey puts the share of DST-opponents in the German population at nearly three quarters (forsa Gesellschaft für Sozialforschung und statistische Analysen mbH, 2015) and there, as well as in many other countries, regular petitions urge parliaments to break with the tradition of changing clocks twice a year.

In addition, and although one can set one's clocks to reading regularly in the newspapers that DST should be abandoned because it is detrimental to school children's performance (e.g., Schmidt, 28.03.2009; Draper, 05.03.2015), there is hardly any scientific evidence on whether the time change affects school performance. To the best of our knowledge, there is only one study (Gaski and Sagarin, 2011) on the long-run impact of the semiannual clock changes and students' performance in the Scholastic Aptitude Test (SAT) in Indiana, USA. The authors report SAT scores in DST-adopting counties to be lower by 16 points (which equals 16% of a standard deviation). Yet, we doubt that the difference the authors find can be explained by the clock change as Gaski and Sagarin (2011) do not account for potential structural differences between counties that might drive the results. Moreover, most US states experience changes of more than 10 points in mean SAT scores in reading and math over time (National Center for Education Statistics, 2013), half of the SAT-taking schools experience a rise or fall in scores by 10 points every year, and about 20% tend to have 20 points higher or lower test scores when compared to the previous year (College Board, 2014).

This chapter is the first to study whether the clock advance induces short-run consequences on students' performance in six European states. We exploit the fact that several countries collected data for the international student assessments *Trends in International Mathematics and Science Study* (TIMSS) and *Progress in International Reading Literacy Study* (PIRLS) during the transition from ST to DST in spring 2011. This approach provides us with a sample of more than 22,000 students. Hypothesizing that elementary school children might suffer from sleep deprivation and a relatively "earlier" school start in the week after the switch to DST, we mimic the underlying basic structure of a regression discontinuity design by investigating whether moving the clock forward by one hour affects students' performance.

This mechanism is backed up by a rich literature on the relationship between sleep and performance, which indicates that cumulative or complete sleep deprivation decreases cognitive test performance (e.g., Astill et al., 2012; Van Dongen et al., 2003; Banks and Dinges, 2007; Goel et al., 2009). Whether the rather mild and short-term disturbances of the circadian system introduced by the clock change are large enough to affect children's school performance significantly is, however, unknown. If so, this would not only make a case for another debate on whether to abandon DST or not, but would also call into question the validity of exams and international student achievement tests timed around the clock change. If the shift into DST does not cause large enough drops in students' performance, this would cast doubt on the trustworthiness of the claim that children suffer measurably from the clock change.

Our results challenge the predominant expectation that the clock change introduces strong and measurable changes in children's school performance. Although we do find small decreases in performance after the clock change in most countries for math and science, these effects are very small in magnitude and not significantly different from zero at neither point of the performance distribution. Moreover, the treatment effects for reading are pointing to the opposite direction and are of similar magnitude, though also not statistically significant. Our results are moreover robust to varying the time window around the clock change and cannot be explained by the young age of the fourth-graders in our main sample.

2.2 Effects of the clock advance

2.2.1 The circadian clock

In each of us ticks a circadian clock that determines when we sleep and when we wake. Daylight serves as a zeitgeber to our inner clock and synchronizes our sleep-wake patterns approximately (circa) to the daily (dian) rotation of the Earth. Our organism is tied to that inasmuch as the hormone melatonin regulates our sleep-wake-cycle by sending us to sleep. When it gets dark, our bodies produce melatonin and we begin to feel sleepy. At dawn, the production is stunted and we awake.

In Europe, this bio-chemical system is disturbed each spring when the clock is set forward by one hour in the very early morning hours of the last

Sunday in March, and our alarm clock conflicts with our inner clock: While school or work still start at, say, 8 a.m. clock time, our bodies continue to follow the light-dark cycle that still lags one hour behind. In the mornings, melatonin levels are still up and we feel sleepy. In the evenings, we have difficulty falling asleep. This deprivation of sleep persists until the DST and the light-dark cycle are synchronized, or in other words, until we have settled the dispute between the alarm and inner clock (Valdez et al., 2003, p. 146).

2.2.2 Sleep, light, and cognitive performance

Both sleep and light are also correlated with cognitive performance. Light does not only affect vision but exerts a direct positive effect on the functioning of the brain and its availability increases cognitive performance (Heschong et al., 2002; Vandewalle et al., 2006, 2009).

The positive association between sleep duration and cognitive test performance of adults is well documented (e.g., Van Dongen et al., 2003; Banks and Dinges, 2007; Goel et al., 2009).[3] A recent meta-analysis shows that also for 5-12 years aged children, sufficient sleep is significantly related to higher cognitive performance, less internalizing (e.g., anxiety, sadness) and externalizing (e.g., aggression, hyperactive behavior) behavioral problems, and, especially, better performance in school (Astill et al., 2012, and references therein). At the same time, children's attention, memory, and intelligence seem to be unaffected by sleep duration (Astill et al., 2012).

Correlational studies draw the picture of a positive relationship between self-reported hours of sleep and grades in middle and high school (consult Wolfson and Carskadon (2003) or Shochat et al. (2014) for a review). Children seem to be sensitive to small or modest changes in sleep duration. In that vein, Vriend et al. (2013) show that reducing habitual sleep duration of 32 children by one hour for four consecutive nights affected children's mood and emotional regulation negatively and decreased their cognitive performance.

2.2.3 Sleep and performance after the clock change

It is not clear, in how far all these processes carry over to the clock change, especially with respect to elementary school children who are the subject of this study.

[3] The performance enhancing effects of sleep even seem to pay off in monetary terms. Instrumenting sleep duration with the local sunset time, Gibson and Shrader (2014) estimate the causal effect of hours slept on wages. Speculating that an earlier sunset drives people to bed earlier, the authors provide evidence that sleeping one hour more each night increases wages by 16%.

First, there is mixed evidence on how long the sleep-wake cycle needs to adapt after the clock change. Results from both early and recent studies indicate that children and adolescents lose between 40 and 50 minutes of sleep following the switch from ST to DST (Reese, 1932; Barnes and Wagner, 2009). Schneider and Randler (2009) report that school children showed a higher daytime sleepiness after the time change. The adaption process to the new regime can take up to several weeks, depending on chronotype and sleep patterns during weekends (Valdez et al., 2003; Schneider and Randler, 2009). In contrast, adjustments to phase delays as encountered when clocks are reset to ST, traveling westwards, or moving from daytime shift work to night shift work are easier and faster (Hauty and Adams, 1965a,b; Lemmer et al., 2002; Niu et al., 2011).

Second, the impact of a single small short-term shift in the circadian clock is only rarely studied and if so with small sample sizes and in an artificial setting. For instance, Burgess et al. (2013) simulate small disturbances of the circadian clock in 11 adults, who reacted with significantly slower reaction times in a Psychomotor Vigilance Test. Monk and Aplin (1980) analyze the performance of 39 adults during the shift from DST to ST, i.e., during the phase delay in fall. After waking under the standard clock time, subjects showed enhanced performance in calculation tests. Yet, the authors cannot separate this effect from the simultaneous effect of a better mood on awakening.

A separate strand of the literature analyzes how delaying school starting times affects educational achievements. Although these studies are rather focused on medium-term outcomes than on the effects of short-term disturbances of the circadian clock, we want to briefly review this literature as delaying school start by one hour mimics the reverse of the shift into DST at least temporarily.

Recently, three economic studies provided quasi-experimental evidence of a later school start time on students' achievements. Although these studies can rely on larger samples and exogenous variation instead of self-reported measures from survey data, the results are, again, mixed.

Edwards (2012) exploits the fact that US middle schools start the school day at different times to reduce the costs of the public transportation system. Using between and within variation, he finds a 2-3 percentage point increase in standardized math and reading test scores when school starts one hour later.

Carrell et al. (2011) show that delaying course start times by 50 minutes increased students' achievements at a US-military post-secondary institution by as much as a one standard deviation increase in teacher quality. The authors use variation from two sources: First, starting times were shifted step-wise from 7:00 to 7:30 and finally to 7:50 a.m. Secondly, some students were randomly allocated to early courses only, others to later courses only, and the rest attended both early and late courses. Note, however, that students allocated to later courses could not use the additional time in the mornings to sleep longer because they were required to attend the early breakfast with their fellow students. Carrell et al. (2011) argue that late-starting students could have taken a nap between breakfast and their first class, thereby getting more sleep and performing better throughout the day. Given that the military institution prohibited napping (p. 78), it is contradictory that additional sleep should be the main driver of higher performance. To us, it seems equally likely that students in late courses achieved higher grades because the empty time-slot allowed them to repeat and, thereby, better remember the course content. This could also explain why the treatment estimates of attending an early class lose their statistical significance once student fixed effects are included.

In contrast to Carrell et al. (2011), Hinrichs (2011) uses longitudinal individual data on the US high school achievement test ACT and exploits, in his main analysis, exogenous variation from a policy change in the US: While Minneapolis and some of its surrounding districts shifted school starting hours backwards, its Twin City St. Paul and surroundings retained the old starting times. The author does not find evidence for the hypothesis that ringing the school bell later increased students' performance.

Heissel and Norris (2015) take a different approach to investigate the effect of school starting times on students' performance. They exploit the differences in the availability of sunlight before school between time zones both in an instrumental variable approach and a geographic regression discontinuity design. Within-comparisons of students who move across the time zones in Florida show that starting schools one hour later increases test scores up to 0.1 standard deviation for pubescent adolescents, whereas the effects are smaller and insignificant for prepubescent children. Their estimates from the regression discontinuity design in Tennessee generally support these findings.

To the best of our knowledge, there is only one study on the relationship between DST and performance of students. Using the variation in DST-regimes between counties of the US State of Indiana, Gaski and Sagarin (2011) identify the long-run effects of DST on county-wide SAT test performance. The authors find test results to be significantly worse in counties that advance and set back their clocks each year when compared to counties sticking to ST permanently.

Note that our approach, outlined in the following section, is different. Gaski and Sagarin (2011) compare long-run average performance in counties that do or do not change their clocks and interpret their results as persistent difference in performance. Their approach comes at the risk of mistakenly interpreting structural differences between counties as causal effects. The authors do, for instance, not control for the proximity to large cities outside Indiana. It seems plausible that the counties close to Chicago, Cincinnati, or Louisville change the clocks to synchronize working times for commuters from Indiana. Worse SAT scores could then, e.g., be due to the reduced time commuting parents and their children spend at home together or a less privileged background. The latter would also explain why the families cannot afford living closer to the city. In contrast to that, our study focuses on short-run effects of the clock change within DST-adapting countries. Exploiting the random allocation of schools to test dates before and after the clock change as a natural experiment allows us to separate the effect of the transition into DST from structural or institutional differences. If the mild disturbance of the inner clock affects sleep patterns so much that performance in the week after the change suffers, we should be able to observe a short-run dip in performance.

2.3 Method

We analyze the shift to DST as a natural experiment to study before-after differences in students' performance. As sleep-wake cycles and human performance are thought to synchronize within about one week after the clock change (Valdez et al., 2003), we restrict the sample to schools tested within one week before and one week after the change to DST.[4]

More specifically, we regress the test score TS_{ijc} of student i in school j and country c on the treatment indicator, DST_{ijc}, a set of controls, $\boldsymbol{x}_{ijc}$, and the constant η_0. α_0 is the coefficient of interest as it captures the effect of the switch to DST on student test scores. We run hierarchical linear models (HLM) with maximum likelihood to account for the nested structure of the data.[5] The error term is, therefore, a composite taking care of the different

4 As we show in the robustness checks, our results are not sensitive to the time restriction.

5 We also estimated ordinary least squares (OLS)-models with standard errors clustered on the highest level, i.e., on the school level in the country-specific regressions or the country level in the pooled sample. The coefficients estimated with OLS were similar. Moreover, all our results are very similar in a typical regression discontinuity design where the running variable is equal to the number of days away from DST.

variance between schools, υ_j, countries, ν_c, and the remaining individual errors, ϵ_{ijc}:

$$TS_{ijc} = \eta_0 + \alpha_0 \cdot DST_{ijc} + \boldsymbol{x}_{ijc}^{'} \cdot \boldsymbol{\beta} + \upsilon_j + \nu_c + \epsilon_{ijc}. \qquad (2.1)$$

As students are only tested once—either before or after the clock change—our research design relies on the identification assumption that the assignment to test dates before (control group) and after (treatment group) the shift to DST was random. The TIMSS and PIRLS testing dates are restricted to a given time span determined by the end-dates of the school year. Within this time window, school coordinators and testing agencies agree on a specific date. Given sufficient capacity on the test agency's side, the students' performance is assessed on that day.[6] The sampling is, therefore, unrelated to regional characteristics (south/west, rural/urban) that might have also driven the test score results.[7] This procedure might, however, open up the possibility of self-selection into treatment and control group. For example, if coordinators of good schools preferred test dates before the shift to DST because they anticipate a dip in their students' performance, we may mistakenly contribute a negative treatment effect to the clock change, while it only captures a generally worse performance of students tested later.

Although we cannot fully rule out that consideration of the clock change mattered when school coordinators proposed a testing date, we consider it unlikely that coordinators were aware of the clock change and its potential harmful effect on their students' performance as dates were scheduled well in advance.[8] Moreover, assessments took place towards the end of the respective school terms, i.e., during a period where schools schedule examination board meetings, field days, or other activities filling the students' and teachers' timetables. Therefore, we expect that it is challenging enough to arrange a test date that fits the students', teachers', and testing agencies' schedule without consideration of the clock change. Apart from that, it is impossible to identify single schools in the data later, reducing any possible incentive

6 Most schools were tested only on one day per study. In 3.57% of TIMSS- and 3.35% of PIRLS-schools, a few students were tested after the clock change, although their school was sampled before the clock change—probably because they were ill during the main testing time and data for the missing students was collected later.

7 A systematic geographical sampling would have introduced the risk of mistaking structural differences or differences in the availability of daylight between eastern and western areas within a country for a performance difference with respect to the DST-shift.

8 According to the National Research Coordinators of the five TIMSS countries in our sample, the majority of schools was first contacted some 6 to 8 months prior to the testing day.

for school coordinators to optimize their students' performance with respect to test time selection, given that they were indeed aware of the clock change date.

We check the plausibility of the identifying assumption by comparing treatment and control group students on variables that might drive test performance. To do this, we test whether covariate means differ statistically significantly between treatment and control group. To account for the fact that very small differences between treatment and control group lead to high values for the t-statistic if the sample size is large, we also calculate the scale-free normalized differences as suggested by Imbens and Wooldridge (2009, p. 24). More specifically, we take the differences in means between covariates before, x_{before}, and after, x_{after}, the treatment and normalize them by their sample standard deviations, using the respective sample variances before, s^2_{before}, and after, s^2_{after}, the treatment:

$$\Delta_x = \frac{x_{after} - x_{before}}{\sqrt{s^2_{before} + s^2_{after}}}. \tag{2.2}$$

Following the authors' rule of thumb, we interpret differences larger than a quarter of a standard deviation as indication of selection bias and sensitivity of linear regression with respect to model specification.

To account for potential differences in performance over the week, e.g., a "blue Monday effect" or exhaustion over the week (Laird, 1925; Guérin et al., 1993), but also to investigate whether the DST-effect fades out over the week, we include control variables for each testing day of the week, day, and its interaction with the treatment indicator:[9]

$$\begin{aligned} TS_{ijc} =& \eta_1 + \alpha_1 \cdot DST_{ijc} + \sum_{d=2}^{5} \gamma_d \cdot day_{id} + \sum_{d=2}^{5} \delta_d \cdot day_{id} \cdot DST_{ijc} \\ & + \boldsymbol{x}'_{ijc} \cdot \boldsymbol{\beta} + \upsilon_j + \nu_c + \epsilon_{ijc}. \end{aligned} \tag{2.3}$$

We use Monday as the reference category. Therefore, α_1 represents the treatment effect for Mondays after the treatment. The marginal effect of the time change on the Tuesday under DST equals then, for instance, the

[9] Please note that we thereby allow the treatment effect to vary non-linearly over days of the week. We also investigated whether imposing a more restrictive functional form, such as quadratic or cubic time trends, change our results. As we did not find evidence of increased fit and our results remained similar, we decided in favor of the specification presented here.

treatment dummy, α_1, plus the coefficient of the treatment interacted with the performance on $d = 2$ after the clock change, δ_2.

2.4 Data

The *International Association for the Evaluation of Educational Achievement* (IEA) has been assessing fourth- and eighth-graders' reproduction, application, and problem solving skills in several areas of math and science since 1995 in the *Trends in International Mathematics and Science Study (TIMSS)*. Moreover, the IEA measures trends in fourth-graders' reading literacy and comprehension every five years in the *Progress in International Reading Literacy Study (PIRLS)*.

We can make use of several fortunate coincidences in the latest currently available waves of 2011 which we use for the following analyses: First, 2011 is the only year for which we can use assessment data for all three testing areas (math, science, and reading) because both TIMSS and PIRLS data were collected. Secondly, while data on the exact date of the testing was not contained in previous waves, this information is available in the 2011 waves. Lastly, as student achievement data were collected in the last months of the respective countries' school terms, the field phases of several countries coincided with the transition into DST. In TIMSS 2011, there was an overlap between fourth-graders' testing dates and the clock change in seven countries. Being especially interested in performance differences on the Monday after the clock change (which was March 28, 2011), we have to exclude the two countries that lack test data on Mondays (Finland and Ireland). After excluding four students who were tested on a Sunday and 309 cases with missings on our covariates, our analytic sample from TIMSS contains 8,813 fourth-graders in 364 schools from Denmark, Lithuania, Norway, Spain, and Sweden.[10] As Denmark did not participate in PIRLS 2011, our PIRLS sample includes Lithuania, Norway, Sweden, and Spain, but also Finland where students' reading performance was assessed on all weekdays. After listwise deletion of 357 cases with missing values, our analytic PIRLS sample sums up to 13,255 fourth-grade students clustered in 508 schools.

TIMSS and PIRLS follow a matrix-sampling approach, meaning that there are many more questions asked in total than answered by a single student in the assessment booklets. Whereas students answer only one booklet, each

[10] We focus on students in grade 4 as the data for eighth-graders do only include two countries (Sweden and Finland) for the respective time period and reading literacy is not assessed in grade 8. We do, however, draw on the eighth-graders sample in our robustness checks.

item is contained in more than one booklet. The IEA uses this overlap to construct an estimate of the achievements in the student population with the help of scaling methods from item-response theory (see Mullis et al. (2009a, p. 123), Mullis et al. (2009b); Yamamoto and Kulick (2012)). To account for the uncertainty introduced by imputing the scores, the IEA provides five *plausible values* of the achievement scores. We retain this uncertainty by using all five plausible values in the following analyses.[11]

Achievement scales range usually from 300 to 700 points. To establish comparability over time and between countries, the IEA scaled achievement test scores in 1995 (TIMSS) and 2001 (PIRLS) to an international mean of 500 and a standard deviation of 100.

Note that we focus on TIMSS for the following short description of our data in order to save space. We provide statistics for our PIRLS sample in the appendix (tables A2.1 and A2.2) and outline only main points in this section.

Table 2.1 gives an overview of the plausible values for the countries in our sample, showing that students perform between the intermediate and the high international benchmark, which are set at 475 and 550 points. On average, students achieve scores of about 509 points in math (S.D.$\approx$72) and 515 points in science (S.D.$\approx$70). Spanish students score lowest and Danish as well as Lithuanian students highest in math. In the science assessment, test scores are highest for Swedish and lowest for Norwegian children. For PIRLS, we find an average of 536 points in reading (S.D.$\approx$68), and that the best readers in our sample are the Finnish students, while the elementary school children in Norway achieve the lowest scores (cf. appendix table A2.1).

Table 2.1 contains further descriptive statistics on our later controls. We include gender and age in months to investigate heterogeneous effects and to account for potential differences between treated and controls. All of our students are in grade 4 and the average student is about 10 years old. Half of the sample is female. The high performing Danish, Finnish, Swedish, and Lithuanian students are, on average, one year older than the lower performing Norwegian and Spanish children.

11 More specifically, we apply Rubin's Rules (Rubin, 1987) to combine adjusted coefficients and standard errors from all plausible values as implemented in the multiple imputation (mi) commands in Stata.

Table 2.1: Descriptive statistics (TIMSS)

	Pooled	Denmark	Lithuania	Norway	Spain	Sweden
Student Performance						
Math	509.14	537.29	537.47	496.54	490.41	505.95
	(71.97)	(68.66)	(71.57)	(68.81)	(69.15)	(67.50)
Science	515.47	528.38	518.35	496.23	515.67	534.84
	(69.90)	(71.01)	(65.32)	(64.13)	(70.93)	(74.84)
Student demographics						
Female	0.50	0.54	0.49	0.51	0.49	0.48
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age (months)	122.70	130.55	128.30	116.59	117.40	128.76
	(7.28)	(4.59)	(4.28)	(3.48)	(5.00)	(3.93)
Test language spoken at home						
always	0.79	0.83	0.83	0.80	0.74	0.78
	(0.41)	(0.37)	(0.37)	(0.40)	(0.44)	(0.41)
sometimes	0.18	0.16	0.16	0.18	0.19	0.21
	(0.39)	(0.37)	(0.36)	(0.39)	(0.39)	(0.41)
never	0.03	0.01	0.01	0.01	0.07	0.01
	(0.16)	(0.08)	(0.10)	(0.12)	(0.26)	(0.11)
Books at home						
$<$ 1 shelf ($<=$10)	0.09	0.09	0.13	0.06	0.09	0.06
	(0.28)	(0.28)	(0.33)	(0.24)	(0.29)	(0.24)
1 shelf (11-25)	0.25	0.27	0.36	0.19	0.26	0.20
	(0.44)	(0.45)	(0.48)	(0.39)	(0.44)	(0.40)
1 bookcase (26-100)	0.35	0.37	0.35	0.36	0.34	0.34
	(0.48)	(0.48)	(0.48)	(0.48)	(0.47)	(0.47)
2 bookcases (101 - 200)	0.17	0.16	0.10	0.20	0.16	0.22
	(0.37)	(0.37)	(0.30)	(0.40)	(0.36)	(0.41)
$>$ 2 bookcases ($>$200)	0.14	0.11	0.07	0.19	0.15	0.18
	(0.35)	(0.31)	(0.25)	(0.39)	(0.36)	(0.38)
Day covariates						
Monday	0.14	0.14	0.10	0.14	0.17	0.15
	(0.35)	(0.35)	(0.30)	(0.35)	(0.37)	(0.36)
Tuesday	0.27	0.17	0.24	0.22	0.35	0.29
	(0.44)	(0.38)	(0.43)	(0.42)	(0.48)	(0.46)
Wednesday	0.30	0.48	0.29	0.34	0.24	0.28
	(0.46)	(0.50)	(0.45)	(0.47)	(0.43)	(0.45)
Thursday	0.20	0.17	0.21	0.23	0.20	0.18
	(0.40)	(0.37)	(0.41)	(0.42)	(0.40)	(0.39)
Friday	0.09	0.04	0.16	0.07	0.05	0.09
	(0.28)	(0.19)	(0.37)	(0.25)	(0.21)	(0.28)
Observations	8813	564	2116	2328	2208	1597

Notes: Own calculations for the pooled sample based on TIMSS 2011. Mean values and standard deviations (in parentheses) of the pooled and country-specific samples. The day covariates indicate the percentage of students tested on that day. We used all five plausible values and applied Rubin's rule (Rubin, 1987) to calculate the appropriate standard deviations of the average student performances.

We include an indicator for whether students wrote the test in the language they speak at home to control for language-related differences in test scores. About 80% do indeed always stick to the test language at home and only 2-3% indicate to never use it at home. To control for the children's socio-economic background by proxy, we add the number of books at home.[12] Most children indicate that their parents have 26-100 books (one bookcase) at home. In 14% of the cases, children report more than two bookcases (more than 200 books) at home. The average number of books at home is relatively high in Sweden, Norway, and Finland, though relatively low in Lithuania.

When turning to test days, the table shows that most students were tested on Tuesdays or Wednesdays. In our TIMSS-sample, 14% of the overall sample was tested on Mondays (table 2.1), thereof 43% before and 57% after the switch to DST. 15% of the PIRLS-students were tested on a Monday (appendix table A2.1), 28% of them under ST and 72% under DST.

As outlined in the previous section, we test for (normalized) differences between students treated before and after the clock change. The results are reported in table 2.2 for TIMSS and appendix table A2.2 for PIRLS. While absolute differences are statistically significantly different from zero for most variables, they show neither a systematic pattern nor are the normalized differences above the critical value of 0.25 suggested by Imbens and Wooldridge (2009, p. 24). Including these covariates in the following regressions controls for slight differences between groups that should not substantially affect our results.[13]

---

12 There are three main reasons why we favor this often used proxy for the educational, social, and economic background of the family. First, the number of books at home is easily comparable across countries (Wößmann, 2004). Second, the predictive power of the books variable with respect to student performance is higher than that of parents' educational background (Wößmann, 2003). Third, while books at home are reported for nearly all students, parents' educational achievement is systematically missing for about one third of the cases in our TIMSS sample and about 12% in our PIRLS sample. Missing cases are a selective sample of students with a low number of books at home.

13 We also investigate differences in other proxies for the students' socio-economic status between "treated" and "untreated" students, e.g., own possessions including books, study desks, or computers. We do not find a systematic pattern within and over countries that would point to a selection of specific students or schools to the treatment or control group. Moreover, our results are very similar after including these variables as additional controls.

Table 2.2: Differences in covariates before and after the treatment (TIMSS)

	Before		After		Before - After		Normalized difference
	Mean	(S.D.)	Mean	(S.D.)	Diff.	(P-value)	
Student demographics							
Female	0.50	(0.50)	0.50	(0.50)	0.00	(0.93)	−.001
Age (months)	123.51	(7.11)	122.03	(7.36)	1.48	(0.00)	−.144
Test language at home							
– always	0.77	(0.42)	0.81	(0.39)	−0.04	(0.00)	0.076
– sometimes	0.21	(0.41)	0.16	(0.37)	0.05	(0.00)	−.090
– never	0.02	(0.15)	0.03	(0.17)	−0.01	(0.11)	0.024
Books at home:							
– less than one shelf (<=10)	0.08	(0.27)	0.09	(0.29)	−0.01	(0.03)	0.032
– one shelf (11-25)	0.24	(0.43)	0.27	(0.44)	−0.02	(0.01)	0.038
– one bookcase (26-100)	0.34	(0.48)	0.36	(0.48)	−0.01	(0.19)	0.020
– two bookcases (101 - 200)	0.17	(0.38)	0.16	(0.37)	0.02	(0.06)	−.029
– more than two bookcases (>200)	0.16	(0.37)	0.13	(0.33)	0.03	(0.00)	−.069
Day of test:							
– Monday	0.13	(0.34)	0.15	(0.35)	−0.01	(0.05)	0.029
– Tuesday	0.20	(0.40)	0.32	(0.47)	−0.12	(0.00)	0.195
– Wednesday	0.37	(0.48)	0.24	(0.43)	0.13	(0.00)	−.203
– Thursday	0.19	(0.39)	0.21	(0.41)	−0.02	(0.00)	0.044
– Friday	0.10	(0.30)	0.07	(0.26)	0.03	(0.00)	−.071
Observations	4000		4813		8813		

Notes: The table reports the mean values and standard deviations of all covariates for the control group (tested before the transition into DST) and the treatment group (tested after the transition). The third column contains the difference in means and the respective p-values from testing the hypothesis that the two means are equal. The last column reports the normalized differences as suggested by Imbens and Wooldridge (2009, p. 24). Calculations based on TIMSS 2011.

2.5 Results

2.5.1 Performance-effects of the clock change in the pooled sample

Table 2.3 reports the effects of the clock change on students' performance in math, science, and reading for the pooled sample of all countries. Note that the students tested in math were also tested in science and vice versa, because both fields were part of the TIMSS study. Most of the TIMSS students did also participate in the PIRLS reading assessment, but not all of them.[14]

Table 2.3: Impact of the clock change on students' performance (pooled sample)

	(1)	(2)	(3)	(4)
	Sample: TIMSS (8,813 observations)			
a) Mathematics				
DST effect	−4.042	−9.131	−3.462	−8.139
	(3.512)	(8.742)	(3.074)	(7.663)
b) Science				
DST effect	−3.892	−10.601	−3.433	−9.439
	(3.444)	(8.586)	(2.909)	(7.293)
	Sample: PIRLS (13,255 observations)			
c) Reading				
DST effect	0.506	8.180	0.309	4.072
	(2.695)	(6.626)	(2.306)	(5.686)
Sociodemographic controls			✓	✓
Days & interactions		✓		✓

Notes: Own calculations for the pooled sample based on TIMSS and PIRLS 2011. Sociodemographic controls: gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always); day and interaction controls: weekday (reference: Monday), weekday×DST. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

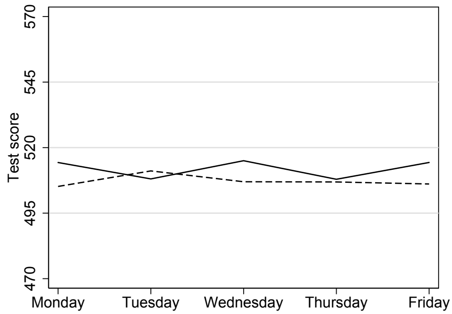
In the HLM specification without covariates (table 2.3, column 1), students scored about 4 points lower in both math and science when tested during

[14] We would have liked to also present within-analyses for students who participated in both TIMSS and PIRLS and completed one study before and the other one after the clock change. Unfortunately, TIMSS and PIRLS were always conducted at consecutive days within the same week. Nevertheless, we verified that the results presented here are not sensitive to the order in which the tests took place.

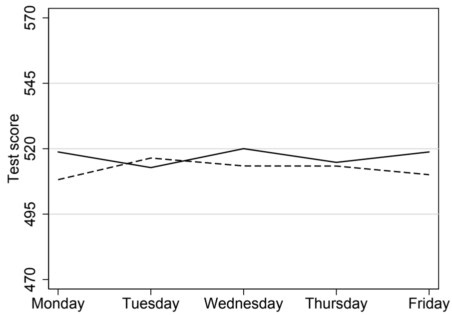
the week after the clock change. Given a standard deviation of about 72 and 70 points, test scores in the week after the clock change drop by roughly 6% of a standard deviation. Yet, these effects are neither substantial in terms of statistical significance nor in terms of magnitude. This gets even clearer when looking at the estimate for reading, indicating that students performed less than 1% of a standard deviation better when sampled after the clock change. Again, the effect is not statistically different from zero.

Two points stand out after including weekday dummies and weekday-treatment interaction terms as of equation 2.3 in column 2. First, students' patterns of performance in the week before the time shift are rather stable (figure 2.1).[15] Second, as the Monday before the time change is our reference category, the treatment coefficient in column 2 shows the reduction in students' test scores at the Monday immediately after the switch to DST. We would expect the treatment effects to be largest for the first day of the week when only one night passed since the clock had been advanced. For both math and science, the coefficients imply that the negative effect on students' test scores is indeed strongest on Mondays after the shift where students might suffer most from sleep deprivation. Contradictory to that, PIRLS students performed slightly better in reading on the Monday after the clock change when compared to the average over the week. Yet, as the standard errors increase by about the same rate, all point estimates remain statistically insignificant. Plotting the average performance levels shows that students' achievement scores decrease only slightly after the clock change as can be seen from the dashed line in figure 2.1 for math (panel a) and science (panel b). For reading, we observe a very small positive effect (panel c). Please keep in mind, however, that these differences are not statistically significant and equal to the magnitude of normal fluctuations over weekdays.

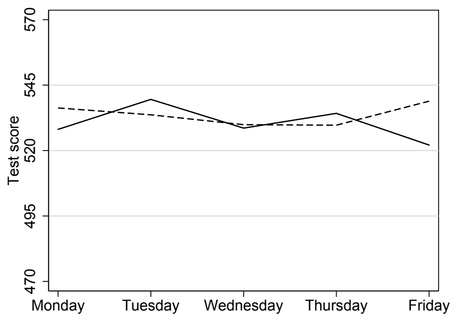
15 To the best of our knowledge, there are only two studies on day-of-week effects in students' performance with contradictory results. Laird (1925) finds students' performance highest on Wednesdays. While Laird (1925) uses a multi-faceted measure of cognitive abilities, Guérin et al. (1993) test 8- to 10-year old girls' attention, mental speed, and visual scanning abilities in a letter cancellation test. Guérin et al. (1993) do not find a pattern for 8-year-olds but do find peaks in performance for the 10-year-old girls on Tuesdays or Fridays.



(a) Math



(b) Science



(c) Reading

Figure 2.1:
Performance before (solid line) and after (dashed line) the clock change over weekdays

Notes: Own calculations based on TIMSS and PIRLS 2011. The ordinate was scaled to a range of roughly 0.5 standard deviations above the average performance in reading and 0.5 standard deviations below the average performance in math.

Adding covariates for gender, age, age squared, books at home, and whether test language is spoken at home instead (column 3) does only slightly decrease the estimates of the treatment effects when compared to column 1, confirming our identification assumption.[16] As given in full detail in the appendix tables A2.3 to A2.5, the signs of all covariates moreover follow expected patterns with female students scoring lower in math and science but higher in reading, and diminishing positive effects of students' age on test scores.[17] In the TIMSS-sample, students who never talk in the test language when at home score at least one third of a standard deviation below those who always speak it. Obviously, this effect is even larger for reading abilities: Children who never use the test language at home score half a standard deviation lower than children who took the test in their mother tongue. The coefficients for books at home also show the expected signs: Students with low socio-economic status and up to 10 books at home score nearly 75% of a standard deviation lower in science than those with a socio-economic status close to the sample average. The respective effects in math and reading are a bit smaller (68% and 63% of a standard deviation).

Considering the full model (table 2.3, column 4), where *DST* needs to be interpreted as the treatment effect on Mondays after the shift, yields slightly different estimates for reading but not for math and science when compared to column 2. In terms of effect sizes, the DST effect on students' performance in math and science is about as large as the respective gender differences, though insignificant. In reading, the DST effect is about three times smaller than the gender difference.

We test whether girls' or boys' performance is more sensitive to sleep deprivation in table 2.4, columns 1-2. We find no indication of gender effects, which is in line with Monk and Aplin (1980).

Children's sleep deprivation might be dependent on their socio-economic background if the latter is correlated with factors that determine the children's sleep duration as well, e.g., parents' weekend work. We report heterogeneous treatment effects by students' confidence with the test language (table 2.4) and the number of books at home (table 2.5).[18] Although students who report a very large number of books on their parents' shelves were most strongly affected on Mondays after the clock change in the TIMSS-sample, this effect

[16] Adding country fixed effects does not affect our results. As we do not have enough degrees of freedom in all testing areas, we report the most parsimonious models only.

[17] Older students achieve better test score results as we already noted on the descriptive level. As higher age is, however, also a sign of grade repetition or late school enrollment due to possible developmental delay of the child, test scores do not proportionally rise with age.

[18] For these and all other following estimations, we only give the treatment effects in order to save space. We display full results in the appendix.

is still far from being significantly different from zero and not backed up by the PIRLS-sample. We also do not find coherent indication of heterogeneous effects when we split the sample by whether the test language is spoken at home.

Table 2.4: Students' performance by gender and by whether test language is spoken at home

	Gender		Test language spoken at home		
	(1) Female	(2) Male	(3) Always	(4) Sometimes	(5) Never
			Sample: TIMSS		
a) Mathematics					
DST effect	−1.969	−3.977	−3.556	−4.861	−4.450
	(3.471)	(3.625)	(3.195)	(4.453)	(12.214)
b) Science					
DST effect	−3.391	−2.919	−3.359	−5.686	−1.980
	(3.276)	(3.248)	(2.936)	(4.961)	(12.302)
Observations	4393	4420	6977	1600	236
			Sample: PIRLS		
c) Reading					
DST effect	−0.303	1.902	0.338	−1.934	8.364
	(2.608)	(2.590)	(2.433)	(3.571)	(9.759)
Observations	6595	6660	10845	2083	327

Notes: Own calculations based on TIMSS and PIRLS 2011. Regressions include the following socioeconomic controls: gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.5.2 Performance-effects of the clock change in the country-specific samples

Table 2.6 presents the treatment effect estimates by countries. It can be seen that the pattern we described previously for the pooled sample is also reflected in most of the country-specific samples. The effects on students' performance are strongest in Norway and Sweden. Norwegian and Swedish students score about one third (29% and 37% respectively) of a standard deviation lower in math on Mondays after the clock change. The equivalent treatment effects for science are approximately 22% of a standard deviation in Norway and a

Table 2.5: Students' performance by books at home

	(1) 0-10	(2) 11-25	(3) 26-100	(4) 101-200	(5) 200+
			Sample: TIMSS		
a) Mathematics					
DST effect	−1.804	−0.623	−3.076	−4.084	−6.001
	(6.212)	(4.235)	(3.609)	(4.528)	(5.858)
b) Science					
DST effect	−3.464	−0.231	−2.981	−3.900	−5.547
	(6.626)	(3.988)	(3.355)	(4.738)	(5.732)
Observations	758	2245	3096	1456	1258
			Sample: PIRLS		
c) Reading					
DST effect	3.066	4.463	−0.255	0.614	−0.388
	(4.687)	(3.056)	(2.975)	(3.539)	(3.855)
Observations	1048	3043	4871	2401	1892

Notes: Own calculations based on TIMSS and PIRLS 2011. Regressions include the following socioeconomic controls: gender (reference: male), age (centered), age (centered, squared), test language spoken at home (reference: always). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

31% of a standard deviation drop in performance in Sweden. But again, none of these effects is significantly different from zero.

Table 2.6: Country-specific impact of the clock change on students' performance

	(1) Denmark	(2) Finland	(3) Lithuania	(4) Norway	(5) Sweden	(6) Spain
			Sample: TIMSS			
a) Mathematics						
DST effect	28.546		1.028	−19.950	−24.616	−4.667
	(25.547)		(20.757)	(15.614)	(15.186)	(16.502)
b) Science						
DST effect	14.143		−1.307	−14.185	−22.154	−6.867
	(29.155)		(21.109)	(12.003)	(14.392)	(14.249)
Observations	564		2116	2328	1597	2208
			Sample: PIRLS			
c) Reading						
DST effect		−10.880	12.194	15.132	−7.796	11.487
		(9.435)	(16.242)	(14.067)	(24.206)	(10.140)
Observations		3502	2125	2267	1773	3588

Notes: Own calculations based on TIMSS and PIRLS 2011. Gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always), weekday (reference: Monday), interaction weekday×DST included as controls. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A word of caution is in order when investigating the test results for Denmark. In Denmark, only 17 control and 64 treatment group students were tested on a Monday. What is more is that no students were tested on the Friday before the phase delay. Therefore, we do not only lack power to identify any significant effect for Mondays and Fridays but get highly imprecise estimates for the treatment effect which might explain the rather large positive though, again, insignificant rise in test scores after the transition into DST.[19]

The results for reading are equally mixed, and estimates show even positive signs in Lithuania, Sweden, and Spain, although all treatment effects are not significantly different from zero.

[19] Moreover, Lithuania sampled only 93% of the international target population, namely those students taught in Lithuanian (Joncas, 2012), whereas all other countries in our sample did not impose a restriction with respect to language of instruction. This sample selection might be one reason why the Lithuanian treatment effects are slightly different.

2.5.3 Extensions and robustness checks

Two weeks before and after the clock change

If the insignificant decreases in performance in the week after the clock change were just noisy deviations from zero, we would expect that our estimates of the test results two weeks after the clock change are also not clear-cut. If, to the contrary, our one-week effects are factual decreases in performance that are not harmful enough to become significant, we would expect one of the following patterns for treatment effects two weeks before and after the time shift: Either the mildly disrupted circadian clocks have synchronized to the light-dark cycle and two-week treatment effects are consistently and remarkably smaller than the one-week effects. Or students' sleep deficits have accumulated because they did not or only slowly adapt their bed hours to the new system. In the latter case, we would expect systematically larger and potentially significantly negative treatment effects.

Considering the longer observation period of two weeks before and after the clock change moves most of our estimates of the average performance in the two weeks after the clock change closer to zero, but only very slightly and not consistently (table 2.7). When we break down the analysis to the five countries in our TIMSS-sample and compare the already discussed performance estimates on Mondays immediately after the clock change (table 2.6) with our new estimates for the 2-week-window (table 2.8), we do also not find a clear trend: The average TIMSS-performance of the two Mondays after the time change is lower in Denmark and Sweden (in the latter does the estimate even turn significantly negative), but higher in Lithuania. This ambiguity is also mirrored in the PIRLS-sample.

Age effects

The fact that we do not find negative effects on performance in school might be due to the young age of the children in our sample. As children need more sleep in general and are rather morning-chronotypes getting sleepy early in the evenings, they might have less trouble falling asleep when sent to bed earlier (Gau and Soong, 2003) and recover fast from a sleep deficit. In line with this reasoning, Gau and Soong (2003) find Taiwanese students in fourth and fifth grade significantly more morning-oriented and with longer hours of nighttime sleep than sixth- to eighth-graders. Likewise, Edwards (2012) reports that older middle school students in the US are more positively affected when delaying school starting times than younger middle school students. At the same time, he finds no treatment effects for elementary

Table 2.7: Impact of the clock change in the pooled sample, 2 weeks before and after

	(1)	(2)	(3)	(4)
	Sample: TIMSS (14,942 observations)			
a) Mathematics				
DST effect	−1.132	−7.164	−0.871	−3.920
	(2.932)	(7.230)	(2.556)	(6.286)
b) Science				
DST effect	−1.757	−10.057	−1.552	−6.260
	(2.947)	(7.281)	(2.467)	(6.073)
	Sample: PIRLS (21,379 observations)			
c) Reading				
DST effect	−1.193	4.324	−1.333	1.524
	(2.408)	(5.671)	(2.080)	(4.974)
Sociodemographic controls			✓	✓
Days & interactions		✓		✓

Notes: Own calculations for the pooled sample based on TIMSS and PIRLS 2011. Sociodemographic controls: gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always); day and interaction controls: weekday (reference: Monday), weekday×DST. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.8: Impact of the clock change by countries, 2 weeks before and after

	(1) Denmark	(2) Finland	(3) Lithuania	(4) Norway	(5) Sweden	(6) Spain
			Sample: TIMSS			
a) Mathematics						
DST effect	13.378		17.113	−9.175	−27.847**	−8.245
	(21.675)		(18.155)	(13.050)	(13.496)	(13.445)
b) Science						
DST effect	3.076		16.361	−9.827	−24.831*	−8.951
	(21.615)		(17.469)	(9.808)	(13.210)	(12.154)
Observations	1213		3755	2971	3522	3481
			Sample: PIRLS			
c) Reading						
DST effect		−8.349	12.633	11.605	1.003	2.373
		(8.754)	(16.182)	(14.519)	(14.226)	(9.844)
Observations		4600	3664	2870	3458	6787

Notes: Own calculations based on TIMSS and PIRLS 2011. Gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always), weekday (reference: Monday), interaction weekday×DST included as controls. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

school students—either because these are younger or because elementary school starts one hour later than middle schools.

Furthermore, if parents anticipate the clock change and slowly familiarize their children to the new time regime in the days before the switch to DST, the sleep deprivation of young children after the time change would be minimal and our assumption of perfect compliance with the treatment would be violated.[20]

Adolescents are not as easily convinced to go to bed one hour earlier. Additionally, chronotypes shift to evening types during puberty, and adolescents have trouble to go to bed early and to rise early in the mornings (Carskadon et al., 1993). Ages 11 and 12 are usually seen as the transition ages from morningness to eveningness (Carskadon et al., 1993). As girls enter puberty earlier, sleep patterns of same-aged students are often found to be gender-specific (Wolfson and Carskadon, 1998; Laberge et al., 2001; Gau and Soong, 2003; Randler, 2011; Heissel and Norris, 2015). More specifically, the relevant transition to puberty affecting sleep patterns (approximately Tanner stage 3, see Campbell et al. (2012)) occurs at median age 11 for girls and at median age 13 for boys as recent evidence from the US suggests (Heissel and Norris, 2015). Heissel and Norris (2015) show, furthermore, that the impact of changing school starting times on students' academic performance increases not only with age but rises markedly for students at ages 11 (girls) and 13 (boys).

Drawing on this literature, we test two additional hypotheses to investigate the robustness of our null results: First, we hypothesize that female students of age 11 or older are significantly more negatively affected by the clock change than younger females, whereas treatment effects for boys above and below age 11 do not differ significantly. Second, we expect that our treatment effects are significantly larger in a sample of eighth-graders than in our fourth-graders sample we have been studying up to now.

To investigate the first hypothesis, we run separate regressions for boys and girls in our fourth-graders sample and interact the DST effect with a dummy equal to one if the student is at least 11 years old. Table 2.9 repeats the results of the regressions without interactions as a benchmark (columns 1, 4) and displays additionally the specifications including interactions (columns 2-3, 5-6). We find no support for the hypothesis that more mature girls are more affected by the clock change: The interactions are small in magnitude and not statistically significantly different from zero.

[20] Note, however, that TIMSS and PIRLS are low-stakes tests where test performance is without consequences for students so that the risk of imperfect compliance is low.

Table 2.9: Fourth graders' performance by gender and age

	Female			Male		
	(1)	(2)	(3)	(4)	(5)	(6)
			Sample: TIMSS			
a) Mathematics						
DST effect	−1.969	−1.205	−1.143	−3.977	−4.814	−4.771
	(3.471)	(3.707)	(3.685)	(3.625)	(3.843)	(3.825)
11 years or older		15.016***	15.226***		10.760**	10.947**
		(4.669)	(4.668)		(4.732)	(4.732)
DST × 11 years or older		−2.774	−2.820		3.147	3.126
		(5.560)	(5.552)		(5.223)	(5.219)
Country Fixed Effects			✓			✓
b) Science						
DST effect	−3.391	−2.585	−2.419	−2.919	−4.809	−4.710
	(3.276)	(3.461)	(3.445)	(3.248)	(3.513)	(3.498)
11 years or older		17.135***	17.434***		7.997*	8.171*
		(4.520)	(4.520)		(4.715)	(4.713)
DST × 11 years or older		−2.842	−2.773		7.437	7.549
		(5.076)	(5.073)		(5.397)	(5.386)
Observations	4393	4393	4393	4420	4420	4420
Country Fixed Effects			✓			✓
			Sample: PIRLS			
c) Reading						
DST effect	−0.303	−0.296	−0.254	1.902	2.668	2.690
	(2.608)	(2.626)	(2.617)	(2.590)	(2.874)	(2.866)
11 years or older		13.471***	13.612***		17.134***	17.227***
		(3.663)	(3.662)		(3.472)	(3.471)
DST × 11 years or older		0.050	0.072		−2.848	−2.787
		(4.029)	(4.025)		(3.949)	(3.947)
Observations	6595	6595	6595	6660	6660	6660
Country Fixed Effects			✓			✓

Notes: Own calculations based on TIMSS and PIRLS 2011. Regressions include the following socioeconomic controls: gender (reference: male), age (centered), age (centered, squared), books at home (reference: one bookcase), test language spoken at home (reference: always). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To shed some light on our second hypothesis, i.e., whether our insignificant results are driven by the fact that the students in our sample are, on average, only 10 years old, we repeated all analyses with 2011-TIMSS-data on eighth-graders who are right in the middle of puberty (about 15 years old). The chronotypes of eighth-graders should have shifted to eveningness so that these students are more at risk to detrimental sleep deficits after the clock change. Only Finland and Sweden sampled 5,591 students in 199 schools within the time period we are interested in. Descriptive statistics show that the Finnish and Swedish students in the sample are very similar and mirror the descriptives of Finnish and Swedish fourth-graders we presented previously.[21]

Table 2.10 contains our estimation results for a time window of one week before and after the clock change, estimated with sociodemographic controls in a pooled model (columns 1-2) and separately for Finland (columns 3-4) and Sweden (columns 5-6). To explore, again, whether older students are more affected and whether the age effect varies with the students' gender, we add interactions between gender, age, and the DST effect in columns 2, 4, and 6.

Comparing the size of the general treatment effects for eighth-graders in math and science as reported in table 2.10 with the effects of fourth-graders in the respective fields (table 2.3, column 3) shows that the effects for eighth-graders are even smaller and not statistically significantly different from zero. The treatment effects for eighth-graders tested on a Monday (not shown) are slightly larger than the fourth-graders effects, though also not statistically significantly different from zero. Moreover, as can be seen from column 2, we find neither convincing evidence that the clock change had a more detrimental effect for older students nor that girls were more negatively affected than boys.

We draw the same conclusions from our country-specific analyses (columns 3-6): We find a slightly positive DST effect in Finland and a negative DST effect in Sweden; both effects are not significantly different from zero (columns 3, 5). Size and significance of the treatment effects reject our hypothesis of more harmful effects for pubertal students.

In sum, all these results do not confirm the hypotheses that DST is more harmful to older students or to more mature girls.

[21] Again, we do not present tables for most of our age effects analyses. All tables are available upon request.

Table 2.10: Impact of the clock change on eighth-graders, 1 week before and after

	Pooled		Finland		Sweden	
	(1)	(2)	(3)	(4)	(5)	(6)
Math						
DST effect	−1.489	−2.945	3.111	1.841	−8.914	−11.087*
	(3.550)	(3.907)	(4.467)	(4.861)	(5.708)	(6.683)
Age	−0.947***	−1.490***	−1.541***	−2.362***	−0.179	−0.617
	(0.199)	(0.348)	(0.265)	(0.520)	(0.289)	(0.446)
Female	−3.652**	−4.764**	−3.904	−4.935	−3.060	−4.251
	(1.790)	(2.318)	(2.459)	(3.145)	(2.962)	(3.248)
Age × DST		0.607		1.096*		0.346
		(0.530)		(0.657)		(0.857)
Female × DST		3.030		2.352		4.439
		(3.469)		(4.108)		(7.083)
Female × Age		0.688		1.132		0.372
		(0.513)		(0.756)		(0.723)
Age × Female × DST		0.003		−0.914		0.925
		(0.822)		(1.052)		(1.214)
Science						
DST effect	−0.842	−2.017	2.562	1.924	−7.554	−8.941
	(3.647)	(4.074)	(4.376)	(4.847)	(6.613)	(7.951)
Age	−0.772***	−1.598***	−1.125***	−2.034***	−0.350	−1.095*
	(0.208)	(0.384)	(0.244)	(0.457)	(0.370)	(0.596)
Female	−5.336***	−6.158**	−3.949*	−4.379	−7.330*	−8.169*
	(1.960)	(2.577)	(2.271)	(2.944)	(3.876)	(4.649)
Age × DST		1.157**		1.107		1.324
		(0.538)		(0.682)		(0.929)
Female × DST		2.261		1.010		2.763
		(4.087)		(4.185)		(9.318)
Female × Age		1.070*		1.359*		0.697
		(0.587)		(0.697)		(0.947)
Age × Female × DST		−0.594		−1.002		−0.099
		(0.789)		(0.987)		(1.406)
Observations	5591	5591	3547	3547	2044	2044

Notes: Own calculations based on TIMSS 2011. Gender (reference: male), age (in months, centered), books at home (reference: one bookcase), test language spoken at home (reference: always) included as controls. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Dependence on students’ position in the distribution of test scores

If students at different ranges of the test score distribution were significantly but reversely affected, our analysis of students’ average performances could mask these effects. Figures 2.2a to 2.2c present the overall distribution of test scores before and after the shift into DST. Especially in math and science, the test score distributions after the clock change are nice copies of the distribution before the clock change, though slightly shifted to the left. These results are very similar to our very small negative regression estimates (table 2.3, column 4). In reading, the differences between both density curves are even smaller and also in line with the regression results. All in all, we do neither find substantial treatment effects nor meaningful differences between the magnitude of treatment effects at any point of the performance distribution.

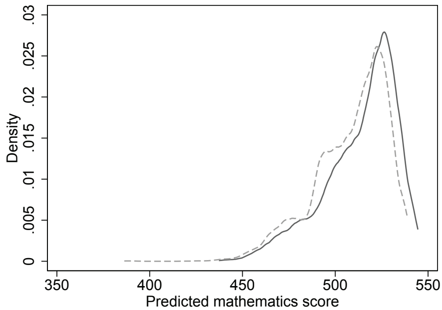
Testing times

Unfortunately, we lack data on the exact times of the assessment, except for Lithuania where the National Research Coordinator provided us with additional information. While the official IEA instruction for schools required to take the tests in the morning hours, we cannot rule out that schools conducted the tests at different times or had different school starting times. As cognitive performance typically increases over the morning hours (Vandewalle et al., 2009), students who took the test later might have performed better. Yet, this would only affect our results if the schools sampled after the clock change had started the assessment systematically later. At least in the Lithuanian case, however, we do not find an indication of assessments being scheduled to later time slots after the clock was advanced by one hour.

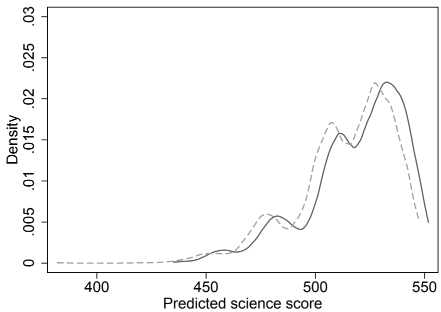
2.6 Discussion

In this chapter, we investigate whether the transition into daylight saving time in spring and a potentially associated sleep loss in the following nights affect students’ performance in international assessment tests. Our findings challenge the prevalent public feeling that the clock change harms school children’s performance.

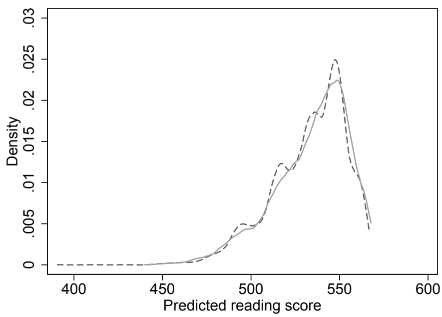
We exploit the fact that schools participating in the large-scale international student assessment studies TIMSS and PIRLS were randomly allocated to testing times before and after the time shift. We mimic the underlying basic structure of a regression discontinuity design by studying differences



(a) Math



(b) Science



(c) Reading

Figure 2.2:
Distribution of performance before (solid line) and after (dashed line) the clock change

Notes: Own calculations for the pooled sample based on all plausible values of TIMSS and PIRLS 2011. The density plots make use of the predicted values from the full model, including all covariates, weekdays and interactions with weekdays.

in performance between students tested before and after the clock change. We do not find elementary school children to be measurably affected by the clock change as our treatment effects are small in magnitude and not statistically significantly different from zero. The estimated treatment effects are most consistent with the hypothesis of a sleep deprivation effect in Norway and Sweden. But even for these countries we fail to find statistically significant effects in the week after the clock change. If these effects, nonetheless, pointed to the Norwegian and Swedish children's higher sensitivity to yet another phase change introduced by DST,[22] we should find large effects for Finland as well. The treatment effects for Finnish students are, however, small, sometimes of positive sign, and not statistically significantly different in all cases. Moreover, for the other countries in our sample and for the testing area reading, we even find positive, yet insignificant, effects.

To investigate several further hypotheses and possible explanations for our findings, we vary the time window around the clock change. Having found no indication of an accumulating or decreasing sleep deficit in elementary school children, we conclude that these young children do not suffer from a harmful sleep deficit. This might be due to the fact that one hour of sleep loss is not enough to disbalance circadian clocks by so much that performance within the following week suffers measurably.

Sleep patterns change with pubertal development and are gender-specific (Carskadon et al., 1993; Gau and Soong, 2003; Heissel and Norris, 2015). If the harmfulness of DST depended on the students' age and maturity in general, we would expect larger treatment effects for more mature girls beyond age 10 and for a sample of older students. To explore these hypothesis, we interact the treatment effect with age and gender. Moreover, we repeat our analyses with a smaller TIMSS sample of 15-year-olds. Having found no indication that treatment effects are larger for pubertal students or higher for more mature girls, we conclude that our null results cannot be explained by the fact that we use a sample of fourth-graders in our main analyses.

All in all, we cannot substantiate the advice that "class and school performance tests should not take place in the first week(s) after the transition into DST" (Schneider and Randler, 2009, p. 1047). We are also cautious to interpret our results as indicative of DST introducing permanent disruptions of the circadian system as Gaski and Sagarin (2011) do for the US because many previous studies show recovery effects in sleep patterns and adaption of the circadian clocks that should not be observed if circadian clocks stay out of

[22] Northern students' circadian rhythms are more often affected by a long, dark winter and a long, bright summer. Previous studies provide evidence that this impairs (mental) health (Rosen et al., 1990; Björkstén et al., 2009).

step. Based on our research, it is fair to say that neither parents nor children nor competence testing agencies (or even policy makers) have reason to worry about allegedly harmful effects of the transition into daylight saving time.

2.7 Appendix

Table A2.1: Descriptive statistics (PIRLS)

	Pooled	Finland	Lithuania	Norway	Spain	Sweden
Student Performance						
Reading	536.35	566.81	536.16	509.48	523.06	539.89
	(67.51)	(63.50)	(64.29)	(61.19)	(65.81)	(66.95)
Student demographics						
Female	0.50	0.49	0.50	0.51	0.49	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age (months)	123.70	129.30	128.35	116.59	117.43	128.87
	(7.28)	(4.19)	(4.26)	(3.51)	(4.93)	(3.95)
Test language at home						
always	0.82	0.89	0.83	0.81	0.77	0.77
	(0.39)	(0.31)	(0.37)	(0.39)	(0.42)	(0.42)
sometimes	0.16	0.10	0.16	0.18	0.17	0.22
	(0.36)	(0.30)	(0.36)	(0.38)	(0.38)	(0.41)
never	0.02	0.01	0.01	0.01	0.06	0.02
	(0.16)	(0.09)	(0.10)	(0.12)	(0.24)	(0.13)
Books at home						
$<$ 1 shelf ($<=$10)	0.08	0.05	0.12	0.06	0.09	0.08
	(0.27)	(0.22)	(0.33)	(0.24)	(0.29)	(0.27)
1 shelf (11-25)	0.23	0.15	0.35	0.19	0.27	0.20
	(0.42)	(0.36)	(0.48)	(0.39)	(0.44)	(0.40)
1 bookcase (26-100)	0.37	0.42	0.35	0.37	0.34	0.34
	(0.48)	(0.49)	(0.48)	(0.48)	(0.47)	(0.47)
2 bookcases (101 - 200)	0.18	0.23	0.10	0.20	0.16	0.20
	(0.39)	(0.42)	(0.30)	(0.40)	(0.37)	(0.40)
$>$ 2 bookcases ($>$200)	0.14	0.15	0.06	0.18	0.14	0.18
	(0.35)	(0.36)	(0.25)	(0.39)	(0.35)	(0.38)
Day covariates						
Monday	0.15	0.13	0.16	0.09	0.26	0.04
	(0.36)	(0.33)	(0.37)	(0.29)	(0.44)	(0.19)
Tuesday	0.27	0.26	0.26	0.27	0.27	0.32
	(0.44)	(0.44)	(0.44)	(0.45)	(0.44)	(0.46)
Wednesday	0.26	0.27	0.20	0.30	0.22	0.36
	(0.44)	(0.45)	(0.40)	(0.46)	(0.41)	(0.48)
Thursday	0.19	0.23	0.20	0.23	0.13	0.17
	(0.39)	(0.42)	(0.40)	(0.42)	(0.34)	(0.38)
Friday	0.13	0.12	0.17	0.11	0.13	0.12
	(0.33)	(0.32)	(0.38)	(0.31)	(0.33)	(0.32)
Observations	13255	3502	2125	2267	3588	1773

Notes: Own calculations for the pooled sample based on PIRLS 2011. Mean values and standard deviations (in parentheses). The day covariates indicate the percentage of students tested on that day. We used all five plausible values and applied Rubin's rule (Rubin, 1987) to calculate the appropriate standard deviations of the average student performance.

Table A2.2: Differences in covariates before and after the treatment (PIRLS)

	Before		After		Before - After		Normalized difference
	Mean	(S.D.)	Mean	(S.D.)	Diff.	(P-value)	
Student demographics							
Female	0.50	(0.50)	0.49	(0.50)	0.01	(0.45)	−.009
Age (months)	124.52	(7.15)	123.00	(7.32)	1.52	(0.00)	−.148
Test language							
always	0.81	(0.39)	0.83	(0.38)	−0.02	(0.02)	0.028
sometimes	0.17	(0.37)	0.15	(0.35)	0.02	(0.00)	−.042
never	0.02	(0.14)	0.03	(0.16)	−0.01	(0.01)	0.030
Books at home							
$<$ 1 shelf ($<=$10)	0.07	(0.26)	0.08	(0.28)	−0.01	(0.08)	0.021
1 shelf (11-25)	0.21	(0.41)	0.25	(0.43)	−0.04	(0.00)	0.067
1 bookcase (26-100)	0.38	(0.48)	0.36	(0.48)	0.02	(0.05)	−.024
2 bookcases (101 - 200)	0.19	(0.39)	0.17	(0.38)	0.02	(0.00)	−.036
$>$ 2 bookcases ($>$200)	0.15	(0.36)	0.14	(0.34)	0.01	(0.05)	−.024
Day of test							
Monday	0.09	(0.29)	0.20	(0.40)	−0.10	(0.00)	0.209
Tuesday	0.32	(0.47)	0.23	(0.42)	0.09	(0.00)	−.139
Wednesday	0.27	(0.44)	0.26	(0.44)	0.01	(0.11)	−.020
Thursday	0.18	(0.39)	0.20	(0.40)	−0.02	(0.02)	0.028
Friday	0.14	(0.34)	0.12	(0.32)	0.02	(0.00)	−.041
Observations	6159		7096		13255		

Notes: The table reports the mean values and standard deviations of all covariates for the control group (tested before the transition into DST) and the treatment group (tested after the transition). The third column contains the difference in means and the respective p-values from testing the hypothesis that the two means are equal. The last column reports the normalized differences as suggested by Imbens and Wooldridge (2009, p. 24). Calculations based on PIRLS 2011. "Test language" indicates whether students speak the test language at home.

Table A2.3: Impact of the clock change on performance in math (pooled sample)

	(1)		(2)		(3)		(4)	
				Fixed Part				
DST effect	−4.04	(3.51)	−9.13	(8.74)	−3.46	(3.07)	−8.14	(7.66)
Day of test								
Tuesday			−6.26	(8.25)			−3.62	(7.32)
Wednesday			0.69	(7.97)			1.87	(7.03)
Thursday			−6.39	(8.20)			−3.73	(7.27)
Friday			0.02	(9.99)			3.55	(8.92)
Interactions								
Tuesday × after			12.18	(10.75)			9.53	(9.56)
Wednesday × after			1.11	(10.71)			2.98	(9.41)
Thursday × after			8.10	(11.01)			7.18	(9.73)
Friday × after			0.92	(13.99)			1.76	(12.29)
Demographics								
Female					−9.71***	(1.51)	−9.69***	(1.51)
Age					0.15	(0.19)	0.15	(0.19)
Age^2					−0.06***	(0.01)	−0.06***	(0.01)
Test language								
sometimes					−10.98***	(2.14)	−10.95***	(2.14)
never					−23.54***	(5.66)	−23.46***	(5.67)
Books at home								
$<=$10					−48.69***	(3.77)	−48.74***	(3.78)
11-25					−21.19***	(1.90)	−21.19***	(1.90)
101-200					9.44***	(2.29)	9.46***	(2.29)
$>$200					10.88***	(2.71)	10.85***	(2.70)
Constant	511.92***	(9.26)	514.31***	(11.14)	529.56***	(9.92)	530.07***	(11.33)
				Random Part				
$\sigma^2_{\upsilon_j}$	392.86***	(261.88)	386.50***	(258.11)	456.33***	(300.87)	447.90***	(295.84)
$\sigma^2_{\nu_c}$	870.46***	(83.63)	862.70***	(83.66)	608.83***	(63.19)	603.51***	(63.35)
$\sigma^2_{\epsilon_{ijc}}$	3973.37***	(73.97)	3972.80***	(74.16)	3692.35***	(67.18)	3691.83***	(67.42)
Observations	8813		8813		8813		8813	
Model F-test	1.32		.49		53.7		33.3	
p-value	.25		.88		.00		.00	

Notes: Own calculations for the pooled sample based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.4: Impact of the clock change on performance in science (pooled sample)

	(1)		(2)		(3)		(4)	
				Fixed Part				
DST effect	−3.89	(3.44)	−10.60	(8.59)	−3.43	(2.91)	−9.44	(7.29)
Day of test								
Tuesday			−5.99	(7.90)			−3.00	(6.87)
Wednesday			1.25	(7.66)			2.37	(6.50)
Thursday			−3.99	(7.97)			−1.16	(6.84)
Friday			−0.01	(9.51)			3.06	(8.20)
Interactions								
Tuesday × after			14.26	(10.72)			10.83	(9.27)
Wednesday × after			4.00	(10.55)			5.75	(8.96)
Thursday × after			9.20	(11.01)			7.56	(9.43)
Friday × after			1.91	(14.07)			3.39	(11.92)
Demographics								
Female					−7.52***	(1.38)	−7.50***	(1.38)
Age					0.24	(0.18)	0.24	(0.18)
Age^2					−0.06***	(0.01)	−0.06***	(0.01)
Test language								
sometimes					−19.60***	(2.10)	−19.56***	(2.10)
never					−30.73***	(4.93)	−30.69***	(4.93)
Books at home								
<=10					−51.79***	(2.55)	−51.87***	(2.55)
11-25					−19.87***	(2.34)	−19.87***	(2.33)
101-200					12.67***	(2.93)	12.70***	(2.93)
>200					14.21***	(2.91)	14.18***	(2.89)
Constant	517.11***	(6.52)	518.77***	(8.81)	534.58***	(6.51)	534.31***	(8.15)
				Random Part				
$\sigma^2_{\upsilon_j}$	173.58***	(121.19)	173.72***	(121.48)	164.29***	(112.89)	164.29***	(112.96)
$\sigma^2_{\nu_c}$	868.10***	(87.57)	863.16***	(88.80)	536.64***	(59.55)	533.96***	(60.94)
$\sigma^2_{\epsilon_{ijc}}$	3884.66***	(82.24)	3883.66***	(82.30)	3534.70***	(72.04)	3533.53***	(72.15)
Observations	8813		8813		8813		8813	
Model F-test	1.28		.517		71.5		44.8	
p-value	.26		.86		.00		.00	

Notes: Own calculations for the pooled sample based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.5: Impact of the clock change on performance in reading (pooled sample)

	(1)		(2)		(3)		(4)	
	Fixed Part							
DST effect	0.51	(2.70)	8.18	(6.63)	0.31	(2.31)	4.07	(5.69)
Day of test								
Tuesday			11.46*	(5.91)			6.29	(5.15)
Wednesday			0.48	(6.15)			−1.49	(5.31)
Thursday			6.12	(6.77)			2.93	(5.77)
Friday			−5.99	(7.05)			−7.13	(6.01)
Interactions								
Tuesday × after			−14.07*	(7.95)			−8.30	(6.86)
Wednesday × after			−6.87	(7.89)			−2.81	(6.73)
Thursday × after			−12.70	(8.81)			−7.36	(7.48)
Friday × after			8.60	(9.65)			8.31	(8.18)
Demographics								
Female					11.69***	(1.07)	11.68***	(1.07)
Age					0.08	(0.13)	0.08	(0.14)
Age^2					−0.07***	(0.01)	−0.07***	(0.01)
Test languagee								
sometimes					−14.59***	(1.76)	−14.56***	(1.76)
never					−34.59***	(3.78)	−34.57***	(3.79)
Books at home								
<=10					−42.55***	(2.28)	−42.54***	(2.28)
11-25					−20.30***	(1.69)	−20.25***	(1.69)
101-200					12.29***	(1.64)	12.23***	(1.64)
>200					13.07***	(2.08)	13.00***	(2.08)
Constant	532.23***	(9.45)	528.09***	(10.71)	538.11***	(8.38)	536.87***	(9.49)
	Random Part							
$\sigma^2_{\upsilon_j}$	427.16**	(275.23)	437.80**	(281.92)	330.71**	(214.23)	336.58**	(217.96)
$\sigma^2_{\nu_c}$	635.66***	(53.35)	621.55***	(51.90)	397.54***	(37.70)	389.02***	(36.62)
$\sigma^2_{\epsilon_{ijc}}$	3548.62***	(67.01)	3546.38***	(66.76)	3222.66***	(54.10)	3221.55***	(53.92)
Observations	13255		13255		13255		13255	
Model F-test	0.04		1.68		113		69.2	
p-value	0.85		0.08		0.00		0.00	

Notes: Own calculations for the pooled sample based on PIRLS 2011. “Test language” indicates whether students speak the test language at home. Age is the centered students’ age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.6: Students' math performance by books at home

	(1) 0-10	(2) 11-25	(3) 26-100	(4) 101-200	(5) 200+
			Fixed Part		
DST effect	−1.804	−0.623	−3.076	−4.084	−6.001
	(6.212)	(4.235)	(3.609)	(4.528)	(5.858)
Student demographics					
Female	−12.504**	−10.493***	−12.246***	−9.502**	−2.659
	(5.702)	(3.050)	(2.459)	(3.782)	(4.498)
Age (months, centered)	−0.325	−0.152	0.436	0.767*	0.418
	(0.533)	(0.370)	(0.344)	(0.431)	(0.554)
Age^2(centered)	−0.038	−0.033	−0.081***	−0.058*	−0.057
	(0.032)	(0.022)	(0.023)	(0.030)	(0.040)
Test language at home					
– sometimes	−8.578	−13.263***	−12.057***	−13.990**	−9.554*
	(7.017)	(3.689)	(3.527)	(5.498)	(5.185)
– never	−8.783	−25.655***	−21.409**	−39.881***	−35.765**
	(12.718)	(8.810)	(8.594)	(12.222)	(17.434)
Constant	472.500***	505.752***	532.322***	544.438***	537.253***
	(13.650)	(11.605)	(8.989)	(10.733)	(9.213)
			Random Part		
$\sigma^2_{\upsilon_j}$	633.584***	603.725***	351.534***	464.932***	262.189***
	(465.891)	(405.745)	(239.585)	(327.337)	(235.418)
$\sigma^2_{\nu_c}$	374.251***	600.191***	531.319***	535.118***	1041.771***
	(203.040)	(101.916)	(93.050)	(114.725)	(220.566)
$\sigma^2_{\epsilon_{ijc}}$	4434.005***	3721.909***	3570.920***	3374.711***	3570.616***
	(372.464)	(137.408)	(102.430)	(167.562)	(198.026)
Observations	758	2245	3096	1456	1258
Model F-test	1.57	5.58	8.78	5.06	1.8
p-value	.15	0.00	0.00	0.00	.10

Notes: Own calculations based on TIMSS 2011. Reference categories used are the following: Test language is always spoken at home, and the student is male. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.7: Students' science performance by books at home

	(1) 0-10	(2) 11-25	(3) 26-100	(4) 101-200	(5) 200+
			Fixed Part		
DST effect	−3.464	−0.231	−2.981	−3.900	−5.547
	(6.626)	(3.988)	(3.355)	(4.738)	(5.732)
Student demographics					
Female	−9.896*	−10.090***	−9.876***	−7.292**	1.699
	(5.321)	(2.893)	(2.618)	(3.320)	(4.060)
Age (months, centered)	−0.364	−0.047	0.641**	1.285***	0.446
	(0.493)	(0.319)	(0.299)	(0.445)	(0.513)
Age^2(centered)	−0.049*	−0.037	−0.080***	−0.067**	−0.054
	(0.029)	(0.023)	(0.025)	(0.032)	(0.038)
Test language at home					
– sometimes	−24.927***	−22.086***	−18.529***	−21.566***	−19.994***
	(6.353)	(3.818)	(3.235)	(4.928)	(4.899)
– never	−20.850	−34.616***	−27.707***	−45.888***	−40.706***
	(13.457)	(8.399)	(8.839)	(12.909)	(15.588)
Constant	477.537***	512.425***	536.594***	552.222***	544.524***
	(10.130)	(6.893)	(5.666)	(6.014)	(8.393)
			Random Part		
$\sigma^2_{\upsilon_j}$	323.847***	154.591***	88.463***	91.727***	215.352***
	(283.227)	(119.601)	(72.132)	(84.800)	(167.073)
$\sigma^2_{\nu_c}$	406.496***	522.330***	466.957***	391.436***	913.834***
	(181.373)	(96.731)	(78.063)	(108.923)	(210.660)
$\sigma^2_{\epsilon_{ijc}}$	4154.278***	3496.890***	3346.028***	3440.415***	3501.204***
	(294.887)	(128.923)	(104.421)	(150.893)	(176.000)
Observations	758	2245	3096	1456	1258
Model F-test	3.89	10.5	11.3	7.59	4.19
p-value	.00	0.00	0.00	0.00	0.00

Notes: Own calculations based on TIMSS 2011. Reference categories used are the following: Test language is always spoken at home, and the student is male. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.8: Students' reading performance by books at home

	(1) 0-10	(2) 11-25	(3) 26-100	(4) 101-200	(5) 200+
			Fixed Part		
DST effect	3.066	4.463	−0.255	0.614	−0.388
	(4.687)	(3.056)	(2.975)	(3.539)	(3.855)
Student demographics					
Female	4.774	5.802***	10.493***	15.875***	19.880***
	(4.425)	(2.225)	(2.028)	(2.621)	(3.103)
Age (months, centered)	−0.618	−0.445*	0.238	1.396***	0.347
	(0.395)	(0.256)	(0.208)	(0.318)	(0.398)
Age^2(centered)	−0.051**	−0.034**	−0.083***	−0.056**	−0.117***
	(0.026)	(0.016)	(0.015)	(0.025)	(0.025)
Test language at home					
− sometimes	−22.337***	−17.405***	−12.654***	−16.461***	−9.367**
	(5.576)	(3.075)	(2.734)	(3.870)	(4.323)
− never	−38.956***	−43.315***	−33.597***	−41.030***	−24.884**
	(9.043)	(6.998)	(7.709)	(8.720)	(11.157)
Constant	493.939***	516.362***	540.384***	551.015***	551.422***
	(10.013)	(8.923)	(7.445)	(6.719)	(9.312)
			Random Part		
$\sigma^2_{\upsilon_j}$	397.711***	361.984***	238.491***	175.009***	356.214***
	(284.590)	(242.106)	(158.843)	(124.227)	(251.569)
$\sigma^2_{\nu_c}$	347.269***	355.208***	414.946***	328.800***	404.640***
	(137.585)	(67.142)	(53.356)	(69.759)	(98.998)
$\sigma^2_{\epsilon_{ijc}}$	3668.163***	3157.304***	3072.398***	3108.141***	3423.860***
	(230.197)	(98.878)	(77.784)	(105.872)	(150.476)
Observations	1048	3043	4871	2401	1892
Model F-test	7.49	14.5	16.7	15.9	13.3
p-value	0.00	0.00	0.00	0.00	0.00

Notes: Own calculations based on PIRLS 2011. Reference categories used are the following: Test language is always spoken at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.9: Students' math performance by gender and by whether test language is spoken at home

	Gender		Test language spoken at home		
	(1) female	(2) male	(3) always	(4) sometimes	(5) never
			Fixed Part		
DST effect	−1.969	−3.977	−3.556	−4.861	−4.450
	(3.471)	(3.625)	(3.195)	(4.453)	(12.214)
Demographics					
Female			−9.053***	−13.265***	−12.473
			(1.719)	(3.497)	(10.849)
Age	0.144	0.198	0.350	−0.567	1.265
	(0.243)	(0.309)	(0.223)	(0.368)	(0.782)
Age^2	−0.048***	−0.055***	−0.054***	−0.044*	0.018
	(0.016)	(0.017)	(0.014)	(0.024)	(0.084)
Books at home					
$<$ 1 shelf ($<=$10)	−52.103***	−50.151***	−50.338***	−49.996***	−43.265***
	(4.332)	(4.686)	(4.360)	(6.076)	(14.598)
1 shelf (11-25)	−21.374***	−22.611***	−20.495***	−24.888***	−22.566*
	(2.628)	(2.691)	(2.077)	(4.442)	(12.225)
2 bookcases (101-200)	11.698***	8.518**	10.504***	10.493*	−5.320
	(2.905)	(3.824)	(2.468)	(6.046)	(15.287)
$>$ 2 bookcases ($>$200)	16.332***	6.343*	11.581***	13.281**	−5.117
	(3.553)	(3.703)	(2.723)	(6.612)	(17.929)
Constant	519.236***	530.766***	528.879***	521.369***	494.728***
	(10.915)	(9.199)	(9.754)	(11.455)	(14.296)
			Random Part		
$\sigma^2_{\upsilon_j}$	547.636***	368.686***	434.414***	541.115***	0.000
	(362.138)	(253.431)	(288.645)	(373.489)	(0.000)
$\sigma^2_{\nu_c}$	575.952***	605.084***	597.647***	444.245***	1592.131***
	(76.291)	(83.316)	(66.573)	(131.136)	(668.232)
$\sigma^2_{\epsilon_{ijc}}$	3484.272***	3896.057***	3640.317***	3919.167***	3634.655***
	(95.922)	(97.238)	(83.412)	(177.383)	(550.371)
Observations	4393	4420	6977	1600	236
Model F-test	35.1	30.8	47.9	18.5	1.61
p-value	0.00	0.00	0.00	0.00	.12

Notes: Own calculations based on TIMSS 2011. Age is the centered students' age in months. Reference categories used are the following: the student reports one bookcase (26-200) of books at home, and the student is male. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.10: Students' science performance by gender and by whether test language is spoken at home

	Gender		Test language spoken at home		
	(1) female	(2) male	(3) always	(4) sometimes	(5) never
			Fixed Part		
DST effect	−3.391	−2.919	−3.359	−5.686	−1.980
	(3.276)	(3.248)	(2.936)	(4.961)	(12.302)
Demographics					
Female			−6.661***	−9.986***	−15.650
			(1.695)	(3.819)	(10.307)
Age	0.094	0.424	0.450**	−0.282	0.605
	(0.247)	(0.259)	(0.193)	(0.458)	(0.866)
Age^2	−0.058***	−0.056***	−0.060***	−0.043*	−0.019
	(0.016)	(0.016)	(0.013)	(0.025)	(0.095)
Books at home					
$<$ 1 shelf ($<=$10)	−54.492***	−54.342***	−50.198***	−59.950***	−48.874***
	(4.145)	(3.418)	(2.965)	(5.864)	(15.582)
1 shelf (11-25)	−20.795***	−20.551***	−18.505***	−26.688***	−23.729*
	(3.052)	(3.366)	(2.510)	(4.395)	(12.785)
2 bookcases (101-200)	14.432***	12.019***	13.219***	14.056**	4.510
	(3.370)	(3.933)	(3.188)	(5.874)	(18.114)
$>$ 2 bookcases ($>$200)	19.678***	9.719**	15.076***	13.880**	3.345
	(3.389)	(3.868)	(3.159)	(5.588)	(19.227)
Constant	527.994***	534.467***	533.705***	519.398***	501.774***
	(7.051)	(6.290)	(6.332)	(8.533)	(15.118)
			Random Part		
$\sigma^2_{\upsilon_j}$	194.037***	126.471***	146.189***	237.683***	0.000
	(137.412)	(92.653)	(101.849)	(173.809)	(0.000)
$\sigma^2_{\nu_c}$	501.946***	526.170***	518.059***	530.045***	1158.402***
	(72.855)	(73.867)	(63.769)	(128.065)	(790.192)
$\sigma^2_{\epsilon_{ijc}}$	3402.029***	3667.406***	3447.463***	3681.465***	4256.135***
	(103.288)	(93.695)	(75.528)	(153.635)	(565.125)
Observations	4393	4420	6977	1600	236
Model F-test	46.4	42.4	51.6	21.3	1.88
p-value	0.00	0.00	0.00	0.00	0.06

Notes: Own calculations based on TIMSS 2011. Age is the centered students' age in months. Reference categories used are the following: the student reports one bookcase (26-200) of books at home, and the student is male. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.11: Students' reading performance by gender and by whether test language is spoken at home

	Gender		Test language spoken at home		
	(1) female	(2) male	(3) always	(4) sometimes	(5) never
			Fixed Part		
DST effect	−0.303	1.902	0.338	−1.934	8.364
	(2.608)	(2.590)	(2.433)	(3.571)	(9.759)
Demographics					
Female			12.414***	8.364***	0.748
			(1.138)	(2.800)	(8.565)
Age	0.114	0.027	0.377**	−0.886***	0.757
	(0.196)	(0.179)	(0.149)	(0.291)	(0.927)
Age^2	−0.058***	−0.071***	−0.067***	−0.058***	−0.014
	(0.013)	(0.012)	(0.010)	(0.018)	(0.065)
Books at home					
$<$ 1 shelf ($<=$10)	−47.256***	−42.840***	−40.846***	−50.983***	−45.519***
	(3.236)	(3.010)	(2.573)	(5.102)	(12.417)
1 shelf (11-25)	−22.700***	−19.432***	−19.038***	−26.676***	−28.506***
	(2.311)	(2.358)	(1.793)	(3.603)	(10.414)
2 bookcases (101-200)	15.653***	9.850***	13.233***	10.271**	9.834
	(2.470)	(2.273)	(1.645)	(4.738)	(13.267)
$>$ 2 bookcases ($>$200)	18.863***	8.922***	13.439***	12.107**	22.871
	(2.837)	(2.882)	(2.107)	(4.832)	(15.138)
Constant	549.682***	538.700***	537.002***	528.775***	499.871***
	(8.795)	(7.953)	(7.905)	(9.326)	(12.246)
			Random Part		
$\sigma^2_{\upsilon_j}$	364.305***	285.232***	291.196***	358.528***	95.196*
	(236.928)	(188.270)	(189.308)	(247.165)	(235.515)
$\sigma^2_{\nu_c}$	357.049***	405.995***	402.682***	361.163***	721.900***
	(42.826)	(47.988)	(42.370)	(74.956)	(472.802)
$\sigma^2_{\epsilon_{ijc}}$	3105.701***	3317.985***	3154.296***	3309.876***	4011.707***
	(67.001)	(70.707)	(53.613)	(158.299)	(541.070)
Observations	6595	6660	10845	2083	327
Model F-test	66.4	56.8	96.8	29.9	3.7
p-value	0.00	0.00	0.00	0.00	0.00

Notes: Own calculations based on PIRLS 2011. Age is the centered students' age in months. Reference categories used are the following: the student reports one bookcase (26-200) of books at home, and the student is male. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.12: Country-specific impact of the clock change on performance in math

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
DST effect	28.55	(25.55)	1.03	(20.76)	−19.95	(15.61)	−24.62	(15.19)	−4.67	(16.50)
Day of test										
Tuesday	79.91***	(26.91)	−8.91	(19.74)	−17.89	(16.37)	−5.26	(9.48)	−3.56	(16.56)
Wednesday	42.07*	(23.98)	24.94	(17.51)	−4.10	(14.46)	−11.52	(10.06)	−17.60	(16.98)
Thursday	57.81**	(25.30)	−15.63	(21.85)	−19.77	(14.65)	0.41	(10.95)	−19.26	(24.09)
Friday	2.16	(24.38)	14.59	(19.04)	1.72	(17.26)	−3.84	(11.97)	−21.39	(18.02)
Interactions										
Tuesday × after	−88.23***	(33.35)	27.06	(25.29)	21.25	(19.63)	15.82	(17.02)	2.47	(20.03)
Wednesday × after	−16.56	(28.31)	−7.79	(24.13)	13.65	(17.78)	24.44	(19.41)	−5.70	(21.73)
Thursday × after	−12.66	(36.14)	21.58	(27.09)	20.18	(18.69)	34.28*	(19.90)	5.52	(26.55)
Friday × after			−1.47	(26.48)	18.78	(31.76)	33.75*	(20.35)		
Demographics										
Female	−3.37	(5.94)	−5.54*	(3.29)	−9.87***	(2.76)	−12.33***	(3.36)	−13.70***	(2.82)
Age	0.27	(1.87)	1.33**	(0.63)	−0.35	(1.23)	1.88**	(0.87)	−2.03***	(0.31)
Age^2	−0.04	(0.10)	−0.08**	(0.04)	−0.17*	(0.10)	−0.11*	(0.06)	−0.15***	(0.02)
Test language										
sometimes	−21.37***	(8.24)	−11.06**	(4.99)	−13.65***	(3.91)	−10.09**	(4.53)	−4.80	(3.93)
never	16.76	(32.30)	−34.95**	(16.61)	−21.69	(13.50)	−20.18	(15.79)	−21.73***	(6.20)

Table A2.12: Continued

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
Books at home										
<=10	−51.95***	(12.96)	−42.64***	(5.58)	−56.87***	(7.18)	−46.35***	(6.97)	−43.30***	(6.35)
11-25	−16.93**	(7.38)	−17.94***	(3.50)	−23.33***	(4.22)	−27.31***	(4.71)	−20.08***	(3.66)
101-200	19.53**	(8.50)	10.80**	(5.18)	8.26*	(4.27)	14.91***	(4.40)	3.04	(4.08)
>200	14.48	(11.12)	3.41	(6.49)	6.08	(4.26)	25.81***	(5.38)	6.94	(5.10)
Constant	503.64***	(25.17)	532.55***	(16.09)	527.59***	(13.80)	514.06***	(9.10)	514.23***	(13.26)
					Random Part					
$\sigma^2_{v_j}$	262.89***	(134.18)	589.06***	(131.07)	491.20***	(105.80)	283.11***	(92.97)	605.17***	(135.05)
$\sigma^2_{\epsilon_{ij}}$	3524.66***	(319.55)	3946.10***	(129.63)	3739.45***	(124.74)	3428.89***	(155.00)	3426.65***	(110.45)
Observations	564		2116		2328		1597		2208	
Model F-test	4.43		7.51		11.9		10		13.6	
p-value	0.00		0.00		0.00		0.00		0.00	

Notes: Own calculations based on TIMSS 2011. “Test language” indicates whether students speak the test language at home. Age is the centered students’ age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.13: Country-specific impact of the clock change on performance in science

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
DST effect	14.14	(29.16)	−1.31	(21.11)	−14.19	(12.00)	−22.15	(14.39)	−6.87	(14.25)
Day of test										
Tuesday	71.92**	(31.16)	−14.03	(20.49)	−9.14	(12.31)	−3.27	(9.97)	−0.96	(14.32)
Wednesday	34.25	(27.85)	25.39	(17.64)	−2.22	(11.07)	−9.45	(10.84)	−11.85	(14.62)
Thursday	39.91	(30.71)	−9.56	(21.49)	−12.09	(12.03)	10.39	(11.47)	−24.05	(21.14)
Friday	11.66	(25.99)	11.17	(18.78)	7.85	(13.38)	−2.17	(13.36)	−18.84	(16.15)
Interactions										
Tuesday × after	−75.12**	(36.65)	34.86	(27.21)	13.55	(14.46)	15.96	(18.14)	1.21	(17.74)
Wednesday × after	−7.16	(32.36)	−5.30	(24.22)	16.28	(13.76)	27.72	(20.31)	−9.14	(17.90)
Thursday × after	1.39	(43.04)	19.15	(26.77)	12.85	(15.66)	26.43	(20.66)	17.23	(24.17)
Friday × after			4.74	(26.16)	12.92	(24.88)	27.72	(23.61)		
Demographics										
Female	0.15	(6.10)	−5.72**	(2.66)	−6.54**	(3.15)	−8.54**	(3.92)	−10.97***	(2.81)
Age	0.77	(1.84)	1.25**	(0.55)	−0.16	(1.16)	2.08**	(1.00)	−1.73***	(0.35)
Age^2	−0.05	(0.10)	−0.08**	(0.03)	−0.17*	(0.09)	−0.14**	(0.07)	−0.14***	(0.02)
Test language										
sometimes	−26.03***	(8.29)	−11.05***	(3.63)	−23.91***	(3.80)	−28.21***	(4.87)	−12.79***	(4.51)
never	−23.81	(37.08)	−21.32	(14.85)	−30.01**	(12.30)	−26.11	(17.19)	−30.89***	(5.83)

Table A2.13: Continued

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
	Fixed Part									
Books at home										
<=10	−48.82***	(12.34)	−41.26***	(4.73)	−62.17***	(6.60)	−57.09***	(9.86)	−48.50***	(5.49)
11-25	−17.93**	(7.13)	−19.81***	(3.65)	−21.67***	(4.32)	−26.44***	(5.43)	−14.22***	(4.70)
101-200	20.46**	(8.52)	10.44	(6.40)	13.25**	(5.55)	14.79***	(5.02)	9.01*	(4.61)
>200	9.35	(10.23)	1.45	(7.00)	12.10**	(4.98)	24.55***	(5.82)	14.81***	(4.50)
Constant	503.27***	(27.28)	514.13***	(16.02)	520.08***	(11.08)	542.45***	(9.45)	537.27***	(12.16)
	Random Part									
$\sigma^2_{\upsilon_j}$	321.43***	(174.13)	586.93***	(130.51)	245.21***	(68.45)	460.49***	(126.41)	447.58***	(117.15)
$\sigma^2_{\epsilon_{ijc}}$	3823.49***	(323.18)	3144.59***	(126.84)	3187.02***	(98.86)	3932.98***	(172.71)	3702.80***	(133.62)
Observations	564		2116		2328		1597		2208	
Model F-test	3.91		8.42		16.8		11.2		14.6	
p-value	.00		.00		.00		.00		.00	

Notes: Own calculations based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.14: Country-specific impact of the clock change on performance in reading

	(1) Finland		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
Fixed Part										
DST effect	−10.88	(9.43)	12.19	(16.24)	15.13	(14.07)	−7.80	(24.21)	11.49	(10.14)
Day of test										
Tuesday	−12.10	(7.73)	14.63	(17.27)	11.82	(9.84)	−1.14	(21.34)	18.34	(11.24)
Wednesday	−6.89	(7.42)	−3.90	(18.11)	9.09	(11.03)	−18.67	(21.49)	6.65	(12.29)
Thursday	−11.33	(8.18)	10.84	(17.63)	15.18	(11.05)	−4.25	(22.55)	−5.16	(15.89)
Friday	−22.04**	(9.33)	9.58	(17.26)	8.37	(12.67)	−25.81	(21.22)	−3.78	(20.79)
Interactions										
Tuesday × after	17.22	(11.79)	−17.45	(20.41)	−20.76	(15.49)	4.18	(25.19)	−20.70	(13.36)
Wednesday × after	11.95	(11.49)	−9.50	(21.12)	−16.30	(15.59)	8.73	(25.79)	−6.82	(14.57)
Thursday × after	7.75	(12.04)	−1.17	(20.99)	−21.48	(15.76)	0.76	(27.68)	−7.24	(17.76)
Friday × after	21.50	(14.99)	6.08	(21.28)	−14.21	(20.37)	8.36	(31.48)	3.70	(22.46)
Demographics										
Female	18.68***	(2.05)	14.04***	(2.65)	12.37***	(2.55)	9.25***	(2.94)	3.02	(2.27)
Age (months, centered)	3.05***	(0.53)	0.90**	(0.45)	0.34	(1.27)	2.75***	(0.68)	−1.86***	(0.33)
Age^2 (centered)	−0.23***	(0.03)	−0.08***	(0.03)	−0.10	(0.08)	−0.21***	(0.05)	−0.14***	(0.02)
Test language										
sometimes	−23.51***	(3.64)	−15.53***	(3.82)	−14.53***	(4.14)	−10.17**	(4.29)	−6.81**	(3.23)
never	−35.79***	(11.42)	−27.96	(17.48)	−20.25*	(10.43)	−19.78	(12.31)	−37.09***	(4.84)

Table A2.14: Continued

	(1) Finland		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
Fixed Part										
Books at home										
<=10	−42.45***	(4.87)	−40.66***	(4.28)	−50.08***	(5.42)	−40.89***	(6.71)	−34.05***	(4.32)
11-25	−21.08***	(3.02)	−15.03***	(3.29)	−25.74***	(4.38)	−23.59***	(4.70)	−17.77***	(2.74)
101-200	16.78***	(2.83)	14.71***	(5.02)	7.74**	(3.66)	10.91**	(4.94)	7.99**	(3.39)
>200	17.96***	(3.12)	7.83	(6.10)	8.06**	(4.08)	20.33***	(4.72)	7.03**	(3.49)
Constant	564.08***	(6.74)	524.08***	(14.61)	508.44***	(10.98)	544.65***	(20.92)	516.51***	(9.81)
Random Part										
$\sigma^2_{\upsilon_j}$	217.05***	(46.88)	461.81***	(107.64)	221.17***	(59.23)	408.21***	(101.34)	479.76***	(83.50)
$\sigma^2_{\epsilon_{ijc}}$	3168.52***	(86.89)	3070.02***	(111.37)	3058.91***	(107.20)	3225.46***	(126.76)	3250.58***	(92.41)
Observations	3502		2125		2267		1773		3588	
Model F-test	31.6		12.3		14		11.1		17	
p-value	0.00		0.00		0.00		0.00		0.00	

Notes: Own calculations based on PIRLS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.15: Impact of the clock change on performance in math (pooled sample, two weeks before and after)

	(1)		(2)		(3)		(4)	
				Fixed Part				
DST effect	−1.13	(2.93)	−7.16	(7.23)	−0.87	(2.56)	−3.92	(6.29)
Day of test								
Tuesday			−8.00	(6.98)			−4.20	(6.15)
Wednesday			−4.40	(6.71)			−1.86	(5.82)
Thursday			−6.15	(6.77)			−2.86	(5.90)
Friday			−0.99	(7.90)			3.67	(6.95)
Interactions								
Tuesday × after			10.35	(8.75)			5.04	(7.70)
Wednesday × after			7.05	(9.06)			5.41	(7.87)
Thursday × after			4.75	(8.94)			1.89	(7.79)
Friday × after			4.55	(10.52)			0.52	(9.17)
Demographics								
Female					−10.15***	(1.28)	−10.15***	(1.28)
Age					0.08	(0.14)	0.08	(0.14)
Age^2					−0.06***	(0.01)	−0.06***	(0.01)
Test language								
sometimes					−10.71***	(1.48)	−10.70***	(1.48)
never					−17.35***	(3.99)	−17.35***	(3.98)

Table A2.15: Continued

	(1)		(2)		(3)		(4)	
				Fixed Part				
Books at home								
<=10					-46.26^{***}	(2.81)	-46.30^{***}	(2.81)
11-25					-22.13^{***}	(1.43)	-22.14^{***}	(1.43)
101-200					10.16^{***}	(1.72)	10.16^{***}	(1.72)
>200					13.48^{***}	(2.17)	13.49^{***}	(2.16)
Constant	510.90^{***}	(9.32)	515.61^{***}	(10.69)	528.68^{***}	(10.08)	530.45^{***}	(11.08)
				Random Part				
$\sigma^2_{\upsilon_j}$	410.91^{***}	(267.31)	408.04^{***}	(265.59)	485.08^{***}	(313.79)	479.66^{***}	(310.51)
$\sigma^2_{\nu_c}$	950.02^{***}	(69.05)	944.55^{***}	(68.85)	648.37^{***}	(50.26)	645.83^{***}	(50.23)
$\sigma^2_{\epsilon_{ijc}}$	3972.90^{***}	(59.96)	3972.64^{***}	(59.88)	3689.62^{***}	(51.36)	3689.04^{***}	(51.26)
Observations	14942		14942		14942		14942	
Model F-test	0.15		0.34		93		58	
p-value	0.7		0.96		0.00		0.00	

Notes: Own calculations for the pooled sample based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.16: Impact of the clock change on performance in science (pooled sample, two weeks before and after)

	(1)		(2)		(3)		(4)	
				Fixed Part				
DST effect	−1.76	(2.95)	−10.06	(7.28)	−1.55	(2.47)	−6.26	(6.07)
Day of test								
Tuesday			−9.15	(6.71)			−5.07	(5.68)
Wednesday			−5.95	(6.66)			−3.30	(5.56)
Thursday			−6.02	(6.95)			−2.26	(5.91)
Friday			−5.87	(7.83)			−0.66	(6.58)
Interactions								
Tuesday × after			12.42	(8.67)			6.45	(7.27)
Wednesday × after			10.12	(9.00)			7.90	(7.53)
Thursday × after			5.52	(9.28)			1.96	(7.87)
Friday × after			10.77	(11.30)			5.50	(9.45)
Demographics								
Female					−8.13***	(1.08)	−8.13***	(1.08)
Age					0.24*	(0.13)	0.24*	(0.13)
Age^2					−0.06***	(0.01)	−0.06***	(0.01)
Test language								
sometimes					−20.63***	(1.92)	−20.62***	(1.93)
never					−29.06***	(4.35)	−29.06***	(4.36)

Table A2.16: Continued

	(1)		(2)		(3)		(4)	
				Fixed Part				
Books at home								
<=10					−50.93***	(2.18)	−50.96***	(2.18)
11-25					−21.58***	(1.82)	−21.58***	(1.82)
101-200					11.54***	(2.17)	11.54***	(2.18)
>200					16.70***	(2.11)	16.69***	(2.11)
Constant	515.49***	(5.83)	521.51***	(7.90)	534.07***	(5.72)	536.89***	(7.26)
				Random Part				
$\sigma^2_{\upsilon_j}$	141.73***	(96.92)	143.88***	(98.30)	134.67***	(91.40)	136.72***	(92.81)
$\sigma^2_{\nu_c}$	1013.46***	(75.57)	1007.48***	(75.31)	611.99***	(50.75)	610.63***	(50.80)
$\sigma^2_{\epsilon_{ijc}}$	4005.43***	(69.77)	4005.18***	(69.81)	3637.40***	(59.74)	3636.74***	(59.80)
Observations	14942		14942		14942		14942	
Model F-test	0.36		0.37		116		73	
p-value	0.55		0.95		0.00		0.00	

Notes: Own calculations for the pooled sample based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.17: Impact of the clock change on performance in reading (pooled sample, two weeks before and after)

	(1)		(2)		(3)		(4)	
				Fixed Part				
DST effect	−1.19	(2.41)	4.32	(5.67)	−1.33	(2.08)	1.52	(4.97)
Day of test								
Tuesday			5.42	(5.17)			2.84	(4.64)
Wednesday			1.89	(5.39)			0.98	(4.76)
Thursday			4.28	(5.73)			2.91	(5.05)
Friday			−1.16	(5.94)			−2.80	(5.20)
Interactionsy								
Tuesday × after			−7.56	(6.75)			−3.89	(5.95)
Wednesday × after			−5.01	(6.55)			−2.88	(5.69)
Thursday × after			−7.84	(7.20)			−4.18	(6.27)
Friday × after			−4.70	(7.73)			−2.18	(6.68)
Demographics								
Female					11.23***	(0.82)	11.22***	(0.82)
Age					−0.00	(0.10)	−0.00	(0.10)
Age^2					−0.07***	(0.01)	−0.07***	(0.01)
Test language								
sometimes					−15.20***	(1.32)	−15.21***	(1.31)
never					−29.77***	(2.76)	−29.81***	(2.76)

Table A2.17: Continued

	(1)		(2)		(3)		(4)	
				Fixed Part				
Books at home								
<=10					−41.43***	(1.85)	−41.44***	(1.85)
11-25					−19.07***	(1.28)	−19.07***	(1.28)
101-200					11.99***	(1.31)	11.98***	(1.31)
>200					14.01***	(1.64)	13.98***	(1.64)
Constant	532.77***	(9.58)	529.99***	(10.50)	538.31***	(8.56)	537.03***	(9.42)
				Random Part				
$\sigma^2_{\upsilon_j}$	441.90***	(283.32)	445.40***	(285.57)	351.09***	(225.84)	353.26***	(227.26)
$\sigma^2_{\nu_c}$	679.45***	(44.33)	675.45***	(44.16)	440.25***	(32.32)	437.64***	(32.33)
$\sigma^2_{\epsilon_{ijc}}$	3559.27***	(50.75)	3558.98***	(50.78)	3247.62***	(42.65)	3247.39***	(42.61)
Observations	21379		21379		21379		21379	
Model F-test	0.25		0.5		175		105	
p-value	0.62		0.87		0.00		0.00	

Notes: Own calculations for the pooled sample based on PIRLS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.18: Country-specific impact of the clock change on performance in math (two weeks before and after)

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
DST effect	13.38	(21.67)	17.11	(18.16)	−9.17	(13.05)	−27.85**	(13.50)	−8.24	(13.44)
Day of test										
Tuesday	29.04	(23.21)	−9.33	(20.71)	−10.11	(12.82)	−9.64	(8.01)	−0.37	(14.93)
Wednesday	15.10	(19.12)	24.03	(18.44)	4.99	(10.99)	−17.68**	(8.41)	−15.13	(15.21)
Thursday	26.25	(21.07)	−16.23	(22.92)	−7.28	(10.77)	−6.04	(8.42)	−10.74	(21.38)
Friday	9.31	(24.44)	13.97	(20.04)	15.81	(12.30)	−8.70	(9.28)	−7.85	(33.35)
Interactions										
Tuesday × after	−53.64*	(28.77)	11.75	(23.03)	13.42	(16.64)	25.70*	(15.21)	0.92	(17.75)
Wednesday × after	−7.31	(26.07)	−24.88	(20.81)	5.48	(15.36)	40.71**	(17.32)	5.51	(18.75)
Thursday × after	−43.19	(26.47)	9.03	(24.95)	7.87	(15.94)	42.94**	(16.96)	6.79	(23.46)
Friday × after	1.21	(31.29)	−16.22	(22.67)	4.78	(28.48)	40.73**	(17.23)	−0.18	(35.82)
Demographics										
Female	−9.30**	(4.49)	−8.64***	(2.55)	−9.29***	(2.37)	−9.61***	(2.25)	−14.28***	(2.49)
Age	1.53	(0.99)	0.98**	(0.46)	−0.67	(1.24)	2.10***	(0.59)	−2.08***	(0.25)
Age^2	−0.13**	(0.06)	−0.08**	(0.03)	−0.16*	(0.08)	−0.18***	(0.04)	−0.15***	(0.02)
Test language										
sometimes	−15.11***	(5.17)	−12.37***	(3.19)	−13.18***	(3.74)	−8.00**	(3.39)	−7.01**	(3.23)
never	−19.74	(17.17)	−29.35*	(17.27)	−13.56	(11.79)	−2.93	(7.99)	−18.95***	(4.94)

Table A2.18: Continued

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
Fixed Part										
Books at home										
<=10	−44.32***	(7.82)	−46.16***	(4.52)	−51.64***	(6.41)	−41.18***	(4.25)	−43.54***	(4.78)
11-25	−18.91***	(5.17)	−20.15***	(2.63)	−23.36***	(4.02)	−26.45***	(3.09)	−21.38***	(2.95)
101-200	14.72**	(5.87)	8.94**	(4.15)	9.67***	(3.71)	13.98***	(3.06)	5.47*	(3.21)
>200	19.83***	(7.32)	6.68	(5.00)	8.10**	(4.11)	25.11***	(3.81)	6.80*	(4.05)
Constant	536.08***	(18.38)	537.97***	(16.74)	514.56***	(11.33)	515.30***	(7.32)	511.24***	(10.99)
Random Part										
$\sigma^2_{\upsilon_j}$	679.60***	(163.65)	667.35***	(113.77)	442.42***	(84.24)	346.42***	(69.31)	749.50***	(133.93)
$\sigma^2_{\epsilon_{ij}}$	3522.03***	(224.13)	4035.83***	(110.55)	3698.88***	(117.94)	3445.93***	(99.10)	3430.67***	(95.91)
Observations	1213		3755		2971		3522		3481	
Model F-test	7.15		13.1		14.3		20.1		18.8	
p-value	0.00		0.00		0.00		0.00		0.00	

Notes: Own calculations based on TIMSS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.19: Country-specific impact of the clock change on performance in science (two weeks before and after)

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
DST effect	3.08	(21.61)	16.36	(17.47)	−9.83	(9.81)	−24.83*	(13.21)	−8.95	(12.15)
Day of test										
Tuesday	22.13	(23.60)	−14.49	(20.81)	−8.77	(9.18)	−9.61	(8.36)	4.52	(13.70)
Wednesday	11.62	(19.99)	24.61	(17.89)	0.73	(8.35)	−17.98**	(8.90)	−7.41	(14.14)
Thursday	13.84	(21.44)	−10.07	(21.83)	−5.57	(8.35)	−2.29	(9.02)	−9.54	(19.29)
Friday	−1.13	(25.42)	10.62	(19.08)	13.33	(9.46)	−11.49	(10.35)	−4.21	(31.72)
Interactions										
Tuesday × after	−41.63	(28.46)	17.40	(23.22)	13.09	(11.88)	22.38	(15.69)	−2.73	(16.05)
Wednesday × after	−3.24	(25.77)	−22.80	(20.48)	13.53	(11.68)	43.13**	(17.22)	−5.07	(16.66)
Thursday × after	−30.74	(26.98)	4.41	(23.81)	6.09	(12.65)	35.03*	(18.05)	8.33	(21.64)
Friday × after	11.40	(31.86)	−11.88	(21.42)	7.28	(22.24)	41.27**	(20.65)	−0.82	(33.90)
Demographics										
Female	−5.17	(4.27)	−8.24***	(2.00)	−5.85**	(2.90)	−7.35***	(2.51)	−11.65***	(2.34)
Age	2.72**	(1.14)	0.98**	(0.39)	−0.45	(1.13)	2.51***	(0.62)	−1.80***	(0.30)
Age^2	−0.17**	(0.07)	−0.08***	(0.03)	−0.16**	(0.08)	−0.19***	(0.05)	−0.13***	(0.02)
Test language										
sometimes	−22.40***	(5.96)	−12.81***	(3.06)	−24.74***	(3.63)	−27.46***	(3.86)	−15.64***	(3.52)
never	−46.43**	(21.52)	−18.95	(14.46)	−23.80**	(10.70)	−18.28	(11.17)	−30.94***	(4.93)

Table A2.19: Continued

	(1) Denmark		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
Books at home										
<=10	−47.51***	(7.56)	−44.34***	(3.74)	−57.85***	(5.67)	−53.18***	(7.65)	−50.32***	(4.57)
11-25	−23.97***	(5.33)	−22.24***	(2.89)	−21.27***	(3.89)	−27.14***	(3.71)	−14.80***	(3.45)
101-200	12.30*	(7.03)	8.74*	(4.60)	12.99**	(5.10)	12.89***	(3.37)	9.61**	(3.89)
>200	16.81**	(7.40)	6.52	(5.08)	12.86***	(4.12)	24.38***	(3.60)	16.34***	(3.50)
Constant	530.91***	(18.61)	518.94***	(15.96)	515.08***	(8.94)	545.77***	(8.08)	532.90***	(10.56)
					Random Part					
$\sigma^2_{\upsilon_j}$	676.41***	(173.61)	605.24***	(100.48)	220.02***	(61.24)	513.15***	(96.77)	630.55***	(110.66)
$\sigma^2_{\epsilon_{ij}}$	3957.30***	(272.18)	3277.21***	(117.51)	3153.04***	(84.61)	4047.62***	(132.82)	3714.94***	(113.84)
Observations	1213		3755		2971		3522		3481	
Model F-test	7.4		15.4		20.2		22.7		21.6	
p-value	0.00		0.00		0.00		0.00		0.00	

Notes: Own calculations based on TIMSS 2011. “Test language” indicates whether students speak the test language at home. Age is the centered students’ age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2.20: Country-specific impact of the clock change on performance in reading (two weeks before and after)

	(1) Finland		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
					Fixed Part					
DST effect	−8.35	(8.75)	12.63	(16.18)	11.61	(14.52)	1.00	(14.23)	2.37	(9.84)
Day of test										
Tuesday	−12.05	(7.72)	14.56	(17.95)	5.32	(9.63)	5.40	(8.22)	13.48	(11.25)
Wednesday	−7.00	(7.40)	−4.44	(18.81)	10.69	(10.43)	1.77	(8.71)	3.62	(12.88)
Thursday	−11.30	(8.15)	10.60	(18.30)	9.80	(10.16)	8.22	(9.18)	−5.71	(14.73)
Friday	−22.14**	(9.32)	9.23	(17.92)	9.74	(11.79)	−0.02	(8.84)	2.75	(18.83)
Interactions										
Tuesday × after	14.30	(10.53)	−16.88	(19.85)	−14.33	(16.12)	−0.93	(15.29)	−14.35	(12.53)
Wednesday × after	4.94	(10.24)	−3.86	(20.46)	−17.90	(15.87)	1.05	(15.41)	−2.22	(14.26)
Thursday × after	9.10	(11.08)	−7.55	(20.15)	−15.92	(15.69)	−8.37	(17.74)	0.77	(15.97)
Friday × after	12.65	(12.92)	−13.84	(20.46)	−14.93	(20.06)	−13.48	(22.15)	−5.80	(19.86)
Demographics										
Female	18.99***	(1.77)	11.10***	(1.96)	11.62***	(2.30)	12.27***	(2.21)	4.31***	(1.56)
Age	2.93***	(0.46)	0.88**	(0.37)	−0.14	(1.21)	2.54***	(0.46)	−1.37***	(0.20)
Age^2	−0.23***	(0.03)	−0.07***	(0.03)	−0.13*	(0.08)	−0.20***	(0.03)	−0.11***	(0.01)
Test language at home										
sometimes	−23.46***	(3.00)	−17.74***	(3.22)	−14.83***	(3.61)	−15.23***	(2.75)	−7.06***	(2.43)
never	−40.29***	(10.81)	−25.36	(15.46)	−22.24**	(9.01)	−11.58	(8.93)	−30.32***	(3.46)

Table A2.20: Continued

	(1) Finland		(2) Lithuania		(3) Norway		(4) Sweden		(5) Spain	
	Fixed Part									
Books at home										
<=10	−42.32***	(4.34)	−42.98***	(3.58)	−48.19***	(5.10)	−39.94***	(4.46)	−35.01***	(3.10)
11-25	−20.68***	(2.73)	−17.05***	(2.41)	−23.86***	(3.72)	−24.51***	(3.30)	−15.57***	(1.97)
101-200	17.54***	(2.60)	8.27**	(3.56)	9.82***	(3.24)	11.52***	(3.43)	8.27***	(2.36)
>200	17.86***	(2.70)	7.98*	(4.67)	10.84***	(3.39)	22.59***	(3.32)	7.98***	(2.73)
Constant	564.43***	(6.61)	527.51***	(15.01)	509.71***	(10.85)	531.07***	(8.20)	519.18***	(9.50)
	Random Part									
$\sigma^2_{\upsilon_j}$	213.45***	(43.70)	508.43***	(91.12)	270.85***	(56.92)	347.44***	(66.48)	587.57***	(67.76)
$\sigma^2_{\epsilon_{ij}}$	3179.48***	(72.87)	3103.38***	(86.85)	3051.40***	(100.98)	3217.25***	(88.66)	3317.29***	(65.28)
Observations	4600		3664		2870		3458		6787	
Model F-test	40.9		18.3		17.6		23.3		28.1	
p-value	0.00		0.00		0.00		0.00		0.00	

Notes: Own calculations based on PIRLS 2011. "Test language" indicates whether students speak the test language at home. Age is the centered students' age in months. Reference categories used are the following: Test language is always spoken at home, the student reports one bookcase (26-200) of books at home, and the student was tested on a Monday before the transition into DST. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 3

Non-Take-Up of Student Financial Aid: A Microsimulation for Germany

Stefanie P. Herber and Michael Kalinowski

3.1 Introduction and background

Imagine you are a student in financial need and the government offers you about EUR 38,000 to finance your studies at the following conditions: Given your earnings five years after finishing your studies are sufficiently high, you have to repay 20% of the present value in small rates over the next 20 years.[1] Would you accept the offer?

From a traditional economic perspective, it seems irrational not to claim this huge amount of money. This chapter shows, however, that about two fifths of the eligible German students turn down the offered means-tested student financial aid amounts, called "BAföG". We draw upon rich household data from the German Socio-Economic Panel Study[2] for the years 2002 through 2013 to calculate individual taxes and net incomes in a detailed microsimulation model. Drawing upon the simulated data, we determine those students eligible to receive BAföG and calculate their student financial aid amounts. Subsequently, we give insights into potential explanations why

1 The numbers are in present values, calculated at an interest rate of 2%, see Grave and Sinning (2014).

2 Socio-Economic Panel (SOEP), data for years 1984-2013, version 30, SOEP, 2015, doi:10.5684/soep.v30.

students, who would receive lucrative amounts of student financial aid if they filed an application, do not take up BAföG.

Studying non-take-up of means-tested student financial aid is important for three main reasons.

First of all, BAföG aims at reducing inequalities in educational opportunities for students from low-income families. Federal need-based aid would miss its targets if its design prevented eligible students from claiming their benefits and consequently endangered their enrollment at university or fostered later drop out. Previous research shows indeed that, also for Germany where studying is relatively inexpensive, financial factors are related to students' lower transition rates to university (Schindler and Reimer, 2010; Hübner, 2012) and the intergenerational educational mobility is low (OECD, 2014, p. 93). Moreover, students who decide in favor of studying but against taking up need-based aid have to spend a considerable time working to earn their living. This is generally associated with a higher likelihood to prolong studying (Avdic and Gartell, 2015), dropping out of higher education without a degree (Triventi (2014) provides a review), and performing worse academically (Stinebrickner and Stinebrickner, 2003; Callender, 2008). Against the background that completing higher education goes hand in hand with a non-trivial monetary return, the "sheepskin effect" (Heckman et al., 2006, e.g.), social inequalities can corroborate even if students make their way to university.

Apart from that, evidence on the existence of non-take-up or its low elasticity with respect to the benefits available would moreover contribute to explaining the low responsiveness of students' university enrollments to higher student financial aid amounts in industrialized countries (Dynarski, 2002; Rubin, 2011; Steiner and Wrohlich, 2012).

Last but not least, our results have consequences for researchers and policy makers wanting to anticipate or evaluate student financial aid reforms. As shown by Wiemers (2015), ignoring non-take-up when considering an increase in social assistance benefits leads to striking overestimation of the fiscal costs and the number of (factual) beneficiaries involved.

We contribute to two separate strands of literature on non-take-up: One large strand of literature investigates non-take-up of social benefits, especially social assistance, unemployment, and pension benefits. This literature builds mainly on a straightforward utility maximization of consumers who take up benefits as long as the expected amounts exceed the anticipated claiming costs (Moffitt, 1983; Blundell et al., 1988; Anderson and Meyer, 1997).[3] Previous

[3] An extensive review of the literature is beyond the scope of this chapter. Currie (2004), Hernanz et al. (2004), and Finn and Goodship (2014) provide comprehensive reviews.

studies found that the benefits amount available as well as the anticipated duration (Anderson and Meyer, 1997) of support increase the probability that people take up benefits. The counterweight to these encouraging factors are barriers especially introduced by high transaction costs associated with the claiming process, such as complex forms (Currie, 2004), but also information gaps (Strauss, 1977), and stigma costs (Weisbrod, 1970; Moffitt, 1983).

The unifying feature of the literature on non-take-up of social benefits is that benefit amounts have to be calculated for those who do not claim the benefits and for whom data on benefits received is naturally unavailable. Explaining non-take-up requires then finding suitable proxy variables for the expected costs and benefits of (not) taking up.

Although we stick, methodologically, to this strand of literature, we combine it with insights from a second, separate, strand concerned with debt-averse behavior and students' under-usage of student financial aid, mostly students loans.

So far, only few papers have investigated non-take-up of means-tested student financial aid. Among the related previous studies, Kofoed (2015) draws upon data from the National Center for Education Statistics. The dataset already contains imputed needs for students who did not file the Free Application for Federal Student Aid (FAFSA) which is essential for applying for most federal student aid programs in the US. He finds that about one fifth of eligible US-students fail to complete the FAFSA. Although a minor percentage of the non-takers receives financial assistance from elsewhere (King, 2006; Kofoed, 2015), students still forgo significant amounts of aid they would have been entitled to (Kofoed, 2015). Bird and Castleman (2014) show that even after having completed the application process once, 20% of eligible first semester Pell Grant recipients do not re-file the FAFSA in the subsequent year.

Existing US-studies do not account for the potential endogeneity likely to arise from omitted variables driving both the levels of means-tested benefits and the decision to claim the benefit. We contribute methodologically to this literature by addressing endogeneity with an instrumental variable regression and a sample selection model. More specifically, we instrument the factual, means-tested benefit amount with the BAföG system's generosity and with an indicator for whether the student is independently funded. The implications of the latter are twofold: On the one hand, students who are independently funded have been working before their higher education enrollments. Accordingly, they are also likely to have lower benefits, ceteris paribus. On the other hand, parents' income is not considered in the means test if students are independently funded. Therefore, the level of benefits is higher, ceteris paribus. Empirically, benefit amounts and being independently funded are highly positively correlated. Our sample selection model relies, by contrast, on the exclusion restriction that students who have completed a vocational training before studying are more likely to earn high incomes when studying and selecting themselves out of the sample of eligible, financially needy students.

We are not aware of any study analyzing systematically why students forgo these substantial aid amounts. Previous studies provide, however, mixed evidence as to whether information constraints and complexity of the claiming process can explain non-take-up of student financial aid (Bettinger et al., 2012; Booij et al., 2012; Herber, 2015), while the results are heavily dependent on the design of the aid scheme.

Non-take-up might, however, be higher if student aid is provided as a loan but students are not inclined to bear the psychological costs of having debts (Field, 2009; Oosterbeek and van den Broek, 2009; Cho et al., 2015).[4] This debt aversion is mainly driven by risk aversion and the fear to be unable to repay the loan, but also by cultural differences (Boatman et al., 2014). Regarding the zero interest loans studied in this chapter, debt averse behavior is possible (and rational) for individuals who are willing to save but lack self-control to prevent overspending of the benefit amounts (Cadena and Keys, 2013).[5]

For the German case which we focus on here, only some descriptive statistics stemming from a broad survey of students indicate possible reasons why students do not file the application for BAföG (Middendorff et al., 2013, p. 312). Unfortunately, the survey data do not allow to distinguish between eligible and ineligible students. Therefore, it is not surprising that the most frequently reported reasons are high incomes of parents' or partners' (80%), high own incomes and assets (30%), and low anticipated benefits (14%). Yet, 25% of the students also name debt aversion as a reason why they did not file an application. Information constraints and perceptions of the complexity of claiming are, however, not questioned.

Our study confirms the previous finding that longer expected duration of benefit receipt and higher benefits are important factors of higher take-up rates. Nevertheless, the elasticity of the level of benefits with respect to the probability not to take up BAföG is rather inelastic with an estimate of -.41.

4 Note however that, contrary to BAföG, most loans are supplementary and not means-tested.

5 Cadena and Keys (2013) exploit that eligible US-students who have to pay for room and board and live off-campus can receive a part of the interest-free Stafford loan payed in cash rather than as a credit to their university account. The authors show that if students regard different assets as nonfungible and lack self-control to limit their expenses to prevent overspending, non-take-up can be a rational reaction to avoid overspending.

Furthermore, our analyses yield very robust evidence that students socialized by East German parents are considerably less likely to turn down the money, controlling for various characteristics of students and their parents. Moreover, in line with findings from behavioral economics, suggesting that students at risk to exert too little self-control to restrict their consumption to necessary expenditures (Thaler and Shefrin, 1981; Cadena and Keys, 2013), we detect debt-averse behavior of students low in self-control and high in impatience. Last, being able to draw upon older siblings' experience in the application process is related to substantially higher probabilities to claim BAföG.

The rest of the chapter proceeds as follows: After giving an overview of the German funding scheme BAföG, we elaborate on potential explanations for non-take-up and suitable proxy variables, drawing upon the literature presented above and with an eye on the restrictions of our data. We define the non-take-up rate and outline the empirical models in section 3.4. A description of the data and the sample follows, before we present results in section 3.6 and robustness checks in section 3.7. We close with the discussion. The appendices provide more detailed information on the official calculation of the BAföG benefits (section 3.9.1), how we simulate these benefits (section 3.9.2), and additional sensitivity analyses for our microsimulation model (section 3.9.3).

3.2 The German BAföG scheme for higher education students

Need-based income-contingent aid is regulated in the Federal Training Assistance Act (*Bundesausbildungsförderungsgesetz*), called "BAföG". BAföG was introduced in 1971 and aims at providing equal educational opportunities for all students, irrespective of their social or financial background. While a special form of BAföG is available under certain conditions for students at (higher) secondary schools, this chapter is concerned only with the most frequent target group of BAföG: students enrolled in higher education.

For students in higher education, funding is generally provided for the standard period of studying and intends to support the costs of living and studying. Being the most common form of financial aid for higher education students in Germany, BAföG supported approximately 647,000 students in 2014 at public expenses of about EUR 2.28 billion (Federal Statistical Office, 2015a, p.32). Based on the recent official data from 2012 (German Bundestag, 2014), 66.7% of all students were formally eligible for BAföG, i.e., they met the prerequisites to apply but might have been rejected if they did not pass the means test. 28% of these formally eligible received funding—this equaled 18.7% of all enrolled students in Germany.

As can be seen from figure 3.1, the funded students' percentages of all formally eligible students (upper line) and of all students (lower line) show an upward trend since 1998. The lines reflect the BAföG reforms of 2001, 2008, and 2010 (see tables A3.10 and A3.11 for details). The reforms increased the relative scope of BAföG by raising basic income allowances and made BAföG relatively more attractive by increasing the available aid amounts. Yet, the BAföG scheme is neither indexed to the development of incomes or assets nor inflation-adjusted so that reforms are rather used as readjustment to higher price and income levels.

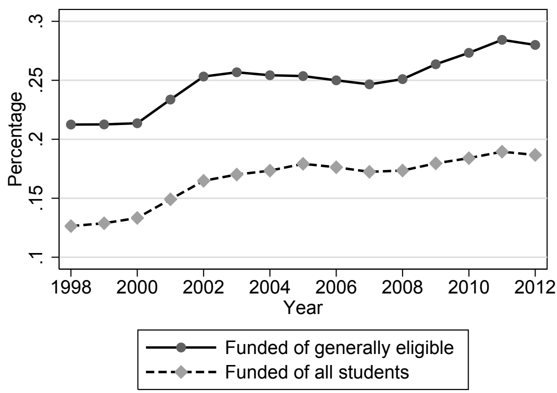


Figure 3.1: Funded students' percentage of the formally eligible and of all students in Germany

Notes: Own figure based on the numbers reported in German Bundestag (2014) and German Bundestag (2010).

BAföG is designed as a grant-loan combination: Half of the amount is generally granted as a subsidy, the other half as a federal zero interest loan. The loan must be repaid within 20 years after a grace period of five years in installments of at least EUR 105 a month. BAföG provides insurance against default risks inasmuch as the repayment is capped at EUR 10,000 and its start can be delayed in case the single, childless claimant's monthly income does not exceed EUR 1,070. The maximum repayment burden for students with very low incomes amounts, therefore, to 9.8%. This burden is in range with proportions of debt usually considered reasonable and bearable (Baum and Schwartz, 2006). Graduates repaying their loan upfront can save monetary amounts up to half of their debts. Grave and Sinning (2014) sum up all direct (grant and loan cap) and indirect subsidies (subsidies of the interest rate). They calculate that students can receive subsidies of up to EUR 30,381, i.e., about 80% of the total BAföG amount (Grave and Sinning, 2014, p. 112).[6]

Before the students' and parents' incomes are considered, students have to meet institutional and personal requirements in order to determine if they are formally eligible at all. The most important requirements (see Appendix 3.9.2 for details) are: Students have to be enrolled in their first degree at higher education institutions, i.e., universities, universities of applied sciences, colleges for professional education, or academies. Furthermore, students must hold German citizenship or have prospects of permanent residency, and, in general, have started their studies before they turn 30 (or 35 for consecutive programs). All students who pass these eligibility checks are formally eligible to receive funding.

Whether *formally* eligible students are also eligible for *positive* funding amounts is then assessed in a means test that proceeds in two steps:

First, the means test takes the students' levels of needs (see table A3.11) as a base value and deducts his or her own economic capabilities. Moreover, the economic capabilities of parents—if they have the legal obligation to support their children—or spouses (or registered partners) are assessed and deducted. If students are older than 30 or have been working for at least five years[7] before enrolling at university, students are independently funded and parents' incomes are not considered for the BAföG calculation. Contrary to the United States' student financial aid system where students' expected expenses resulting from visiting a specific school are imposed, BAföG uses fixed amounts based on the students' living situation. Therefore, students who are not living at the parents' home, have children, or have to cover social security contributions themselves are considered to have additional financial needs which are addressed by (fixed) additions to the basic need levels. Until autumn 2016, the maximum BAföG amount offered to a student who lives outside of the parents' home, has no children but has to pay own social security contributions equals EUR 670. Consequently, the maximum

[6] The maximal subsidy cited here is based on the maximum monthly benefits of EUR 670, a repayment of EUR 105 a month, starting after the grace period, and given an interest rate of 2%. The upfront repayment implies another implicit subsidy of the interest rate, though upfront payment is not worthwhile for high BAföG amounts (Grave and Sinning, 2014, p. 113).

[7] These five years of working experience may include having completed vocational training of up to three years prior to studying.

BAföG amount corresponds roughly to the minimum subsistence level of a single person (German Bundestag, 2015, p. 8). Parents are required to support their offspring up to this maximum rate if the means test results in lower BAföG amounts. The maximum BAföG amount granted reduces to EUR 495 if the student is still living at home.

Second, own income and assets, but also the spouse's, partner's, or parents' income exceeding the respective levels of allowances are subtracted from these general lump-sum amounts, see section 3.9.1 for details. While students' current incomes and assets are relevant, the parents' or spouse's or partner's incomes as of the second last year's tax assessment enter the means test. Students can, however, request that their parents' or spouse's/partner's last year's or current income is used if this is considerably lower than the second last year's income.

The student can generally earn own income from a minor employment paying up to approximately EUR 400 a month without any deductions (see section 3.9.1). Higher earnings are subject to social insurance contributions, personal income tax, and require the student to opt out of the non-contributory dependents' co-insurance, so that most of the students work in jobs that usually pay EUR 400 at a maximum (Middendorff et al., 2013, p. 395).

After accounting for the students' own and familial financial situation, the remaining amount is automatically cashed as a monthly upfront payment to the students' bank account. We refer to all students whose remaining funding amount is positive as "eligible" in the following. In 2014, the average per person per month funding amount (based on the average of the monthly expenditures and assuming that students were funded all year round) was EUR 448; 38% of the funded students received the maximum amounts (Federal Statistical Office, 2015a, p. 32).

3.3 Potential explanations for non-take up of BAföG

From a traditional economic perspective, the student is liquidity constrained, i.e., cannot borrow on the capital market because she cannot offer a collateral for human capital investments. She faces a problem of intertemporal choice where she decides whether or not to take-up BAföG. Given this choice, she maximizes utility from the study and repayment period. In the study period, she can consume both her own income and BAföG or invest it at the capital market to save at the market interest rate. After graduation, the student is constrained by her current income and the repayment of the interest-free loan.

The availability of BAföG during the study period relaxes her budget constraint by allowing her to borrow. Moreover, the subsidies shift her budget constraint outwards so that she can reach a higher indifference curve as long as her preferences are (weakly) monotone and non-satiated. It would, therefore, be rational for the student to accept the money. Even if she does neither want to spend nor invest BAföG at the capital market, she should keep the money at home and pay back the (not inflation-adjusted) loan component some years later.

There might be various reasons for the (seemingly) irrational non-take-up of BAföG. From a rational choice perspective, we can model take-up as the student weighing claiming costs against benefits as has been widely done in the literature analyzing the non-take-up of other social benefits (Blundell et al., 1988; Anderson and Meyer, 1997; Riphahn, 2001; Whelan, 2010; Bruckmeier and Wiemers, 2012, e.g.). Unfortunately, available data sets lack direct measures of the determinants of non-take-up. We discuss suitable proxies and the hypotheses we can investigate with the data at hand in the following.

3.3.1 Utility from claiming BAföG

Previous studies have identified that both the degree and duration of needs influence the utility derived from social benefits positively (Moffitt, 1983; Anderson and Meyer, 1997; Hernanz et al., 2004, e.g.). Accordingly, the probability not to claim BAföG should be higher if students are in higher semesters and closer to the completion of their studies, i.e., the expected duration of the receipt of BAföG is lower. In line with previous research, we proxy the degree of needs by the level of individual, means-tested benefits which result from our simulation. We expect that higher benefits decrease the probability to turn down BAföG, ceteris paribus and that the students will take up BAföG as long as the level of benefits exceeds the claiming costs.

As the students' factual costs of living are not accounted for by the BAföG calculation—apart from a rent subsidy if living outside the parents' home—, we include further proxies associated with the students' factual level of needs. Student financial aid addresses a very homogeneous group of mainly childless, unmarried persons that is similar with respect to age, previous education, and current living situation. Moreover, the BAföG calculation already takes into account contextual factors such as the students' and parents' or partners' living situation and financial capabilities, so that we can restrict our proxies to the individual level. We add a dummy for whether students still live at home because this may decrease their financial need over and above its consideration of the students' place of living in the BAföG calculation. Furthermore, we include an indicator for whether the student lives in East Germany where rents[8]—and therefore need, controlling for parents' income—are lower. To control for differences in living costs but also differences in availability (and accessibility) of minor employment, we also include a dummy for whether the student is living in an urban or rural area.

Compared to the expected family contribution in the US, the German law expects parents to support their dependent children with the amount of their incomes exceeding the respective thresholds (for more details see section 3.9.1). Therefore, we implicitly control for the parents' transfers to their children when keeping the amount of benefits constant. Yet, the official BAföG calculation takes parents' incomes in the second-last year as a default, unless students request using the current, lower, incomes. For that reason, very high current incomes might be associated with higher transfers to the offspring not reflected in the BAföG amount. Consequently, we also add the log of parents' monthly current gross labor income in 2007-EUR.[9]

3.3.2 Disutility from claiming BAföG

Studies investigating social assistance benefits (Riphahn, 2001; Whelan, 2010; Bruckmeier and Wiemers, 2012, e.g.) usually decompose claiming costs into information and stigma costs. Nevertheless, we doubt for three reasons that BAföG involves a social stigma comparable to that possibly felt by persons dependent on social assistance: College is seen as an investment in aspirant future labor market participants. The fact that students do not work (enough to fully finance themselves) is a productive and voluntary "joblessness" because they study full-time and are expected to contribute taxes on their later high incomes after finishing their studies. Moreover, the main calculation basis falls off the person who applies and receives aid so that the reasons for being eligible cannot be attributed to one party. Lastly, the BAföG status cannot be easily inferred from just knowing that someone

8 See Federal Institute for Research on Building, Urban Affairs and Spatial Development (2013), p. 3.

9 We are able to separate the level of benefits and the parents' monthly labor income because the BAföG calculation uses a special, non-deflated income measure. Owing to extensive means-testing and imposition of complex allowances and exemptions, labor income and BAföG benefits are non-linearly related. We report further robustness checks on parents' transfers in section 3.7.2.

is studying. The identification as being poor is, however, a necessary feature of external stigma costs (Weisbrod, 1970).

Different preferences about the welfare state

Nevertheless, we hypothesize that the preferences and perceptions of the welfare state might be different for students socialized in families living in the former socialist German Democratic Republic (GDR) before 1989. Alesina and Fuchs-Schündeln (2007) show that socialism increased the East German's approval of redistribution and provision of social services. While the authors expect the large differences in preferences to prevail for one to two generations (20–40 years) after reunification, i.e., for the sample we consider here, others have shown that differences in social behavior are even more persistent (Brosig-Koch et al., 2011; Heineck and Süssmuth, 2013). Moreover, a recent report demonstrates that East Germans have stronger preferences for high levels of social security and equality and more frequently agree that the state is responsible to achieve these goals (DESTATIS et al., 2013, p. 370ff).

Therefore, we hypothesize that East German families are more likely to regard it as the state's responsibility to provide student financial aid. They should, consequently, find it more natural to take up the assistance they are eligible for than students without an East German background. If this were the case, students with parents living in the East before 1989 should show higher take up rates than similar children to West German parents socialized in an environment more focused on individual responsibility.

To investigate this hypothesis, we include a dummy for whether at least one parent[10] was living in East Germany in 1989 and refer to this variable as "East German background" interchangeably.

Information constraints and complexity of claiming

Students must be aware of the existence of federal aid, be able to understand the aid scheme and file the application. A lack of knowledge and high complexity of claiming the benefits, by contrast, increases claiming costs. A large strand of the literature casts doubt on the assumption of perfectly informed students (Bettinger et al., 2012; Loyalka et al., 2013; Herber, 2015), emphasizes the complexity of federal aid applications (Dynarski and Scott-Clayton, 2006;

[10] In more than 98% of these cases, both parents were living together either in East or West Germany.

Dynarski and Wiederspan, 2012), and shows that information deficits drive non-take-up of other social benefits (Coady et al., 2013).

Our expectations of the role of information constraints and the complexity of claiming aid for the German case are ambiguous: One the one hand, BAföG is the only broad federal student aid scheme and administrated by the student service departments of the universities which makes BAföG a well-known funding source. Moreover, calculators to approximate the prospective benefits are available online (e.g., www.bafoeg-info.de or www.bafoeg-rechner.de/Rechner). On the other hand, students and their parents perceive the 170 questions of the BAföG application forms as confusing and hard to understand; the average time to file the application amounts to 4.5–5.5 hours (Bundeskanzleramt and Nationaler Normenkontrollrat, 2010, p. 41). Apart from that, students might have flawed expectations about their eligibility because the calculation of benefits and the means test are also very complex. In this regard, students might not even consider the possibility that they are eligible, especially if their parents' current labor incomes are high and they are unaware of the fact that the BAföG calculation uses parents' incomes two years ago.

To shed light on the competing mechanisms, we include an indicator for the parents' current labor income, arguing that a higher current labor income decreases not only the perceived level of needs as described in the last section but contributes to the misconception of eligibility. Families with higher current income should, therefore, show a lower probability to take up BAföG if high labor income and high misconception of benefits are correlated, over and above the fact that the need for additional resources is lower.

Moreover, we include an indicator for parents' college degree, assuming that parents with a college degree are, ceteris paribus, better informed about higher education, show higher levels of financial literacy, and might have more resources to assist their children in filing the complex forms. The relationship between non-take-up of BAföG and parents' college degree should, consequently, be negative if a lack of information is important.

In the same line of reasoning, we control for whether students can draw upon the assistance of older siblings who claimed BAföG themselves and are, therefore, well acquainted with filing the forms.

Finally, different groups might lack awareness of the attractiveness of BAföG or the student financial aid system in general. First of all, migrants might suffer from (parents') language barriers or little (parental) knowledge about German student financial aid, making them less likely to file the application. Furthermore and contrary to the positive relationship between East German background and take-up described above, East Germans might equally well show higher non-take-up rates because they have gained less institutional experience with BAföG which was established in West Germany. They might moreover have trouble to file the application because East Germans still lag behind with respect to financial literacy (Bucher-Koenen and Lamla, 2014). If information gaps were more important than different welfare preferences, we would expect a higher non-take-up probability of students with East German background. The existence and direction of the overall effect of the East German background variable is, consequently, unclear.

Parents' experience with public transfers

If East German families or families with a migration background are more likely to be in contact with the public administration, for example, because they receive other welfare benefits already or because they need to file applications for work and residence permits, an economies of scales argument moderates the mechanisms described above: A closer contact to administration officers or receipt of other welfare benefits implies economies of scale when getting informed and filing the applications for BAföG (Dorsett and Heady, 1991).

At the same time, parents' experiences with receiving public benefits may also capture a part of the intergenerational persistence of welfare receipt ("welfare trap"): It might be more socially acceptable for students to claim BAföG if they grew up in a family that received welfare benefits (see, for example, Black and Devereux (2011, p. 1530f) for a review).

To control for these mechanisms, we include a variable for whether someone in the parents' household received public transfers (except maternity benefits and student financial aid) in the previous year. Lacking data on parents' complete welfare receipt histories, we cannot disentangle to which extent our coefficient captures a short-run scale effect or some part of a long-run preference.[11] As both mechanisms point to the same direction, we can, however, hypothesize that parents' (successful) experience with filing forms decreases the likelihood that students reject BAföG if they are eligible.

Time inconsistent preferences, self-control, and debt aversion

Above, we have implicitly assumed a constant exponential discount function resulting in dynamically consistent preferences. Or, in other words, the

[11] In our case, the scale effects argument seems more plausible, however, because we have to restrict parents' welfare receipt to a single year, resulting usually in a downward biased degree of inter-generational persistence in welfare receipt (Page, 2004).

student’s time preferences when deciding about whether or not to take up the aid amount equal those when deciding how to shift consumption between periods. Allowing for hyperbolic discounting relaxes this assumption and can create settings in which consumers wanted to behave patiently in the long-run but are tempted by the immediate gratification of the moment and choose impatiently (Berns et al., 2007, and references cited therein). While impulsivity is the contrary of self-control and associated with impulsive and impatient behavior (Duckworth and Kern, 2011, p. 259),“Self-control refers to the capacity for altering one’s own responses, especially to bring them into line with standards such as ideals, values, morals, and social expectations, and to support the pursuit of long-term goals” (Baumeister et al., 2007, p. 351). Low self-control involves the susceptibility to succumb to impulses, a lack of thinking before acting, not finishing boring or difficult tasks, and striving for exiting, possibly dangerous, activities (Whiteside and Lynam, 2001).

Anticipating their own difficulty to spend the borrowed money reasonably as to limit unnecessary debt—or even anticipating that it might be tough to restrict themselves to pay back the loan after graduating—, sophisticated students might abstain from borrowing completely.

Following the “Economic Theory of Self-Control” (Thaler and Shefrin, 1981), we can think of the student being composed of two selves, one of the selves acting as a far-sighted planner and one as a myopic (low self-control) doer. The far-sighted planner might want to save a part of the benefits not necessarily needed to repay the loan faster. Foreseeing that they will not be able to save because they succumb to their impulses, students might rationally choose a “debt ethic” completely prohibiting borrowing (Thaler and Shefrin, 1981, p. 397). This debt aversion is then not at all irrational but “the logical conclusion of the desire to precommit one’s future economic activity” (Strotz, 1955, p. 173). Indeed, Cadena (2008) and Keys (2009) show theoretically and empirically that, if a sophisticated student is sufficiently impatient and her discount function is quasi-hyperbolic, she rejects an interest-free loan offer in order to limit her own overspending during the study period.

Consequently, we expect present-biased sophisticated students low in self-control not to take out the money and spend it carelessly but rather to show debt-averse behavior and turn down the aid offer completely. As we discuss in more detail in section 3.5, we add two self-reported indicators of low self-control/high impulsivity and impatience and their interaction to our model to test for the existence of the effects elaborated on above. We expect that students are more likely to reject BAföG if they are high both in impulsivity and impatience.

Because the estimated impact of time preferences significantly depends on whether risk aversion is allowed for or not (Andersen et al., 2008), we also control for willingness to take risk, although we do not expect to find an independent effect of risk aversion due to the specific design of the BAföG scheme.[12]

3.4 Method

3.4.1 Definition of non-take-up

Defining a non-take-up rate as the percentage of students who do not take up the benefits available, although they are eligible, requires data on whether the student receives the benefits or not. As eligibility for BAföG is unobservable, eligibility and the respective funding amounts the student would have received had she claimed the benefits must be determined in our microsimulation model.

Four situations can arise when we compare take-up and eligibility: 1., Students simulated as being eligible report funding (take-up), 2., students simulated as eligible do not report funding (non-take-up), 3., students simulated as ineligible report funding (misclassified), 4., students simulated as ineligible do not report funding.

We are mainly interested in why eligible students do or do not claim (cases 1 and 2). Let E denote the number of students simulated as eligible to receive BAföG and let T denote the number of those students who report funding in our data. Let upper bars of these variables represent the contrary, i.e., ineligible $\overline{E}$ and no take-up of the benefit reported $\overline{T}$. The non-take-up rate (NTU) is then defined as the percentage of those who report not to take up the benefits though eligible, $(\overline{T} \mid E)$, to all eligible:

$$NTU = \frac{E - (T \mid E)}{E} = \frac{(\overline{T} \mid E)}{E}. \tag{3.1}$$

While NTU exploits the first two cases arising from our microsimulation model, we can confidently discard the fourth case as ineligible, non-claiming students are of no interest to us.

12 We moreover tested whether our results were affected by omitted variable bias of personality traits that are also strongly associated with self-control (Whiteside and Lynam, 2001). As adding personality traits increases neither fit nor changes our results remarkably, we decided for the more parsimonious models in the following.

Even with high-quality data, it is possible that we classify students as ineligible although they are in fact eligible (case 3). This happens when incomplete or erroneous survey information results in measurement errors. Other than that, students might be classified erroneously as eligible by the public authorities, the administrative process and the students filling in the forms also not being devoid of errors. We use the number of misclassified students to calculate the beta error rate. The beta error rate is defined as the percentage of the students classified as ineligible but reporting benefit receipt $(T \mid \overline{E})$, divided by the sum of all who report to take up the benefits:

$$\beta = \frac{(T \mid \overline{E})}{T}. \tag{3.2}$$

The beta error rate is often seen as a measure of quality of the simulation. This is somewhat misleading because a very detailed eligibility check and a precise calculation of the benefits with the data at hand (potentially containing measurement error) increase the beta error rate (Frick and Groh-Samberg, 2007). Nevertheless, we follow Bargain et al. (2012) and interpret NTU as the upper bound of the non-take-up rate because it ignores those students classified as ineligible by our simulation and calculate a lower bound of the NTU that subsumes misclassified cases under the eligible cases:

$$NTU_L = \frac{(\overline{T} \mid E)}{E + (T \mid \overline{E})}. \tag{3.3}$$

3.4.2 Baseline specification

We can model take up of eligible students in a standard binary choice model where the latent non-take up of BAföG is equal to one if the utility from claiming is larger than the claiming costs (or the utility from non-take-up) and equal to zero otherwise (Moffitt, 1983; Blundell et al., 1988). In our baseline specification, we run a straightforward pooled Probit model and regress our dependent variable NTU on the controls discussed above plus time dummies, age, and gender of the student. We use cluster-robust standard errors to account for the fact that the similarity between observations of a single individual over time is higher than the similarity of observations between different individuals.[13]

[13] In addition to the models presented in the following, we also ran various panel data models. Although the results were mostly identical, we decided in favor of cross-sectional analyses because of the small sample size, the fact that we observe students only twice on average, and the resultant low within and between variations.

3.4.3 Endogeneity of the benefits amount

As students' labor income is deducted from their respective needs, students can influence their level of benefits by earning more or less. If unobserved variables like ability or motivation drive both the level of benefits by higher or lower earnings as well as the decision to file the complex application for BAföG, endogeneity of the level of benefits might bias our estimates. Although incentives to increase own incomes above the threshold of maximum allowances are low, we want to investigate the possibility that endogeneity of the level of benefits affects our results. Thus, we estimate a pooled instrumental variable (IV) Probit model with the structural equation

$$NTU^* = \boldsymbol{z}_1\boldsymbol{\delta}_1 + \alpha b + u_1, \tag{3.4}$$

$$NTU = 1[NTU^* > 0], \tag{3.5}$$

and the first stage for the level of benefits

$$b = \boldsymbol{z}_1\boldsymbol{\delta}_{12} + \boldsymbol{z}_2\boldsymbol{\delta}_{22} + u_2 = \boldsymbol{z}\boldsymbol{\delta}_2 + u_2. \tag{3.6}$$

We assume a bivariate normal distribution of the errors u_1, u_2, independence between the errors and the explanatory variables $\boldsymbol{z}$ (which includes our vector of instruments $\boldsymbol{z_2}$), and normality of our reduced form. If u_1 and u_2 are correlated, our baseline specification suffers from endogeneity. As $u_1|u_2 = \rho u_2 + \epsilon$ and $E(\epsilon|u_2) = 0$, we can formally test whether the benefits level b is exogenous by testing $H_0 : \rho = 0$. We estimate the set of equations by conditional maximum likelihood with clustered standard errors.

As a reference point, we also run a linear two-stage least-squares regression (TSLS) because TSLS requires less distributional assumptions, e.g., errors need not be multivariate normal. Because TSLS ignores the fact that NTU is binary, we again calculate heteroskedasticity-robust standard errors, accounting for the clustered nature and inherent heteroskedasticity of our pooled data.

Similar to McGarry (1996), Whelan (2010), Bruckmeier and Wiemers (2012), and Wiemers (2015), we instrument the level of benefits by the generosity of the system, i.e., the maximum amount of benefits available. Contrary to previous studies on the take-up of social assistance, we can calculate *individual* exogenous maximum benefit amounts because we can exploit the fact that students' benefits do not only depend on their own, endogenous incomes but also on exogenous other features, such as parents' income or family situation. Individual exogenous maximum benefits are more powerful than general maximum amounts: Individual amounts exploit both variation between students due to different exogenous characteristics but also within students over time because of changes in the parents' exogenous characteristics or reforms of the BAföG scheme.

We calculate individual maximum benefit amounts as follows: We take the maximum level of individual needs as a base value by assuming that the student is not living with her parents and receives the maximum rent subsidy. We keep all other factors that determine the student's needs (e.g., whether the student has to pay health insurance herself because she is older than 25 years or has own children) at their observed values as these are arguably not endogenous. From this sum, we deduct only the parents' or the spouse's allowable incomes but not the student's own income or assets. The resulting maximum amounts are, of course, highly correlated with the factual amounts students receive but should, apart from that, not directly drive whether the student claims the money or not.

Our second instrument is an indicator for whether the student is independently funded. The relevance of this instrument exploits the fact that benefit levels and being independently funded are highly correlated: Independently funded students have had the possibility to accumulate higher incomes and assets likely to be deducted from the BAföG funding amounts.[14] Yet, as the parents' income is not deducted, the direction of the effect of being independently funded on the expected level of benefits is, a priori, ambiguous. Exogeneity of the instrument requires that the students' funding states do not directly explain why they accept or reject the money if their income and assets are low enough to yield positive funding amounts.

3.4.4 Selection on eligibility

A last issue we address here is the possibility that students may self-select out of the sample by earning so much that they lose their eligibility to positive funding amounts. Ineligible students are not considered by the non-take-up rate defined above. If sample-selection was relevant, instrumental variable techniques could not account for endogeneity introduced by dropping out of the sample.

Self-selection is a cause of concern as the decision to work and drop out is very likely to be non-random, and the same factors driving this decision might also be correlated with the take-up of benefits. Picking up on the example delineated above, the unobserved motivation and ability of students

[14] The incomes reported by independent students in our sample are about 50% higher than the incomes reported by dependent students.

might simultaneously determine the probability to earn very high additional incomes and the likelihood to successfully file the BAföG application, whereas the direction of this bias is a priori ambiguous. In the example discussed in the last section, the respective level of benefits is simply reduced by the additional income. Here, students' incomes lead to a complete loss of eligibility.

To take into account the incidental truncation caused by the endogenous choice of students' own incomes and assets, we specify a pooled Heckman-type binary response model (Van de Ven and van Praag, 1981):[15]

$$NTU = 1[\boldsymbol{x}_1\boldsymbol{\beta}_1 + \alpha b + u_1 > 0] \tag{3.7}$$

$$y_2 = 1[\boldsymbol{x}\boldsymbol{\delta}_2 + \alpha b + u_2 > 0], \tag{3.8}$$

where b represents, again, the level of benefits. The explanatory variables $\boldsymbol{x}_1$ are a subset of $\boldsymbol{x}$, the cluster-robust errors (u_1, u_2) are independent of $\boldsymbol{x}$ and normally distributed with a mean of zero, a variance of one, and $corr(u_1, u_2) = \rho$. Equation (3.7) is the regression equation with NTU being the binary non-take-up of student financial aid equal to one if the eligible students do not take up their benefits and equal to zero if they do take up. The selection equation is represented by equation (3.8). y_2 is an indicator equal to one if the student's income and assets are below the *individual* threshold of eligibility and equal to zero if the student's income and assets are above the threshold so that she loses eligibility. The non-take-up decision NTU is only observed if $y_2 = 1$, i.e., if the student's income and assets are below their individual thresholds.

To calculate students' individual thresholds, we take the sample of students fulfilling the formal eligibility criteria, including parents' or spouses' incomes, but irrespective of the students' own incomes and assets. We calculate the threshold as the maximum amount a specific student can earn and hold as assets before her simulated benefit amount drops to zero and leads to her self-selection out of the sample. If this drop-out is systematically related to u_1, the estimates of β_1 might be inconsistent.

To identify our system of equations by more than functional form alone, we need at least one variable that is in $\boldsymbol{x}$ but not in $\boldsymbol{x}_1$. As our exclusion restriction, we use a dummy indicating whether the student completed any form of vocational training before studying. Having completed vocational training proxies labor market experience and implies a higher likelihood to

15 Previous to our study, Kayser and Frick (2001) and Frick and Groh-Samberg (2007) used a Heckman-type approach to correct for sample selection into non-take-up of social assistance. Wilde and Kubis (2005) address the issue of sample selection in a simultaneous equation model.

have a job and to earn high incomes while studying. We have to assume that having completed vocational training influences the take-up decision only via the income-channel but does not directly explain (non-)take-up.

3.5 Data and variable construction

Our microsimulation, see section 3.9.2 for details, is based on the Socio-Economic Panel (SOEP v30)[16], which is a representative micro data source for Germany and includes detailed information on household and individual characteristics, as well as extensive information on income (Wagner et al., 2007).

The BAföG calculation was subject to several substantial structural changes between 2001 and 2002, e.g., the unification of needs over Germany and changes in the regulation on additional need amounts, making the system before and after 2001 difficult to compare. Therefore, we restrict our analyses to the waves between 2002 and the most recent wave of 2013. Because we calculate BAföG benefits on an annual basis and according to the law applicable in that year, changes in the BAföG regulation induced by reforms between 2002 and 2013 are taken care of by our microsimulation model.

On the one hand, microsimulation requires high quality data on income and household composition. Analyzing the factors of non-take-up at the same time requires, on the other hand, also suitable proxy variables to be constructed from survey scales. Although the SOEP is generally well-suited for the purpose of microsimulation, not all questions to construct the proxies previously discussed are available for each and every year as we outline in the following.

3.5.1 Constructing the sample and variables

To construct our sample, we proceed in three steps. We keep all students, 1., surveyed between 2002 and 2013, 2., formally eligible for BAföG but not receiving any different student financial aid amounts and, 3., for whom we have enough information to perform the means test and simulate BAföG amounts.

For the last step, we require information on the student's complete family, i.e., parents, siblings, and the student's partners if married or in a registered

[16] Socio-Economic Panel (SOEP), data for years 1984-2013, version 30, SOEP, 2015, doi:10.5684/soep.v30.

partnership. Yet, full information on the parents' incomes[17] is only available for students raised in families drawn as a part of the SOEP—and where parents therefore answer the survey—, but not for cases where students have been drawn as a separate SOEP household after moving out. In order to keep the maximum number of cases for our descriptive analyses, we check whether the student is independently funded or whether the parents died, both cases implying that the parents' income is not relevant for the assessment of eligibility. In these cases, we can keep the student in the sample, although parents' income information is unavailable.

This procedure leaves us with a sample size of 2,827 cases formally eligible to receive BAföG and where enough information on parents' income and living situation is available. Among the formally eligible, about 28% reported to receive BAföG. 53% of all formally eligible cases do not receive BAföG in the SOEP and are also deemed ineligible for positive founding by our simulation. 22% both claim BAföG as reported in the data and are simulated as eligible. 18% are eligible as of our simulation but do not claim the benefits. About 6% of all theoretically eligible observations are beta error observations allegedly claiming benefits but failing eligibility in our simulation.

Some part of this simulation error may be explained by the fact that the SOEP contained only an aggregate measure for all forms of student financial aid through 2006. Consequently, we cannot distinguish between receivers of merit-based aid and those of need-based aid through 2006. Yet, less than 1% of all German students received merit-based aid at this time (Federal Ministry of Education and Research, 2014a). Therefore, this lack in distinction between BAföG and other aid should not be substantive. Accordingly, neither does the simulation quality differ significantly before and after 2007, nor does restricting the sample to the survey years of 2007–2013 affect our results much as we show in the robustness checks later (see section 3.7.3).

For most of the following descriptive analyses, we focus on the group of students simulated as eligible, irrespective of whether they claim BAföG or not, i.e., 1,315 observations. With respect to the sample used for our multivariate analyses, we face the issue that not all of the covariates needed in order to address the possible mechanisms as intended above are available for all years. Moreover, information on parents never questioned by the SOEP could not always be generated from the students' answers. The sample used for our multivariate analysis is, therefore, smaller (i.e., 986 observations).

---

[17] The SOEP provides readily imputed income measures so that we do not lose cases due to item non-response.

In order to prevent a loss of too many observations, we combine responses by parents and information by children about their parents to construct parental background information.

More specifically, we use parents' answers to the question "Where did you live in 1989?" to derive students' East or West German background. If at least one parent indicates to have lived in the East during the fall of the wall, we set the East German background dummy to one and to zero otherwise. The answer to this question is missing only if parents have never been part of the SOEP or were already dead at the time the question was asked. To prevent systematic missings of these cases, we fill the East German background dummy with information on the students' own place of living in 1989 for students already born before 1989.

We face the same issue for the parents' educational degrees. After exploiting the parents' direct information on educational degrees, we substitute missings by using the childrens' information on parents' educational degrees, which is also available if the parents have never been surveyed.

Our indicator for whether the parental household received public transfers in the previous year is, however, unavailable if parents are not part of the SOEP. Accordingly, we can only replace missings as 0 if we know that both parents were already dead last year. All these missings due to the student being sampled as a new SOEP household and the parents never having been surveyed are, however, not systematically related to the factors of non-take-up.

We use survey measures to assess the students' time and risk preferences, all of them measured on a 11-point scale from 0 "not at all" to 10 "very much". The survey questions are worded as follows:

- Impulsivity: "Do you generally think things over for a long time before acting—in other words, are you not impulsive at all? Or do you generally act without thinking things over a long time–in other words, are you very impulsive?"
- Impatience:[18] "Are you generally an impatient person, or someone who always shows great patience?"
- Willingness to take risk: "Are you generally willing to take risks, or do you try to avoid risks?"

Data on impulsivity and impatience were collected only in 2008 and 2013, data on willingness to take risks were collected in 2006 and between

[18] This item was originally reversely coded with 0 representing "very impatient" and 10 "very patient". We reverse the scale to harmonize it with our other measures.

2008–2013 so that we have to assume stability of the concepts over time.[19] Mainly due to the fact that not all eligible students participated in one of the waves where these scales were questioned, our sample is reduced to 986 observations. Yet, again, we see no reason why the year when the student was part of the sample should be systematically related to her non-take-up-behavior. We take the upper quartiles of our impulsivity and impatience scales to construct indicators of high impatience and high impulsivity.

3.5.2 Descriptives

Table 3.1 gives an overview over the weighted analytic sample in general (column 1) and by whether students forgo funding (column 2) or not (column 3). We stick to discussing overall averages, highlighting striking differences by non-take-up in the following.

On average, students are eligible for EUR 314 a month, and, surprisingly, the amount left on the table is only EUR 36 lower on average than the amount taken. Students in our sample are about 23 years old and about half of them is female. Migrants (18% of our sample) are significantly more likely to forgo the benefits (weighted t-test $p < 0.05$). Moreover, we can differentiate between scholarships and BAföG for three quarters of the sample and this percentage does not differ significantly by whether students turn down BAföG or not ($p > 0.1$). Most of those who take up live outside their parents' home and in an urban area, whereas non-takers are much more likely to still live at their parents' home and in rural areas. 17% of the students currently live in East Germany. As can be seen from the numbers of working hours, students who do not take up BAföG work considerably more hours ($p < 0.01$) to support their living.

Remarkably however, students who take out the money do not come from families who are strikingly worse off financially, though non-takers are somewhat less likely to come from a family where at least one parent holds a college degree.

19 The concept of self-control is generally regarded as being stable over the course of life (Gottfredson and Hirschi, 1990; Arneklev et al., 2006) and recent evidence on the longitudinal stability of time preferences elicited in an experimental set-up shows that individual time preferences are also stable for most individuals (Meier and Sprenger, 2015). Harrison et al. (2005) find no significant changes in risk aversion when assessed 6 months later.

Table 3.1: Descriptive statistics by whether students take up BAföG or not

	All		Non-take-up		Take-up	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Simulated BAfoeG amount$^{\natural}$	3.14	(1.39)	2.93	(1.35)	3.29	(1.40)
Age of Individual	23.20	(2.26)	23.05	(2.03)	23.30	(2.41)
Female	0.46	(0.50)	0.43	(0.50)	0.49	(0.50)
Student has direct migration background	0.18	(0.39)	0.22	(0.41)	0.16	(0.36)
Scholarship/BAfoeG can be separated	0.73	(0.44)	0.71	(0.46)	0.75	(0.43)
Living situation controls						
Student living in urban area	0.75	(0.43)	0.83	(0.37)	0.70	(0.46)
Student living at parents' home	0.67	(0.47)	0.80	(0.40)	0.58	(0.49)
Student lives in East Germany	0.17	(0.38)	0.14	(0.35)	0.19	(0.39)
Annual hours worked	197.79	(362.79)	234.16	(404.03)	171.88	(328.17)
Parent and sibling controls						
Parents' current gross labor income$^{\natural}$	31.56	(25.13)	31.06	(19.99)	31.91	(28.24)
At least one parent holds college degree	0.40	(0.49)	0.35	(0.48)	0.43	(0.50)
Parents received social transfers	0.17	(0.38)	0.14	(0.35)	0.20	(0.40)
East German background	0.31	(0.46)	0.21	(0.41)	0.38	(0.49)
Older sibling claimed BAfoeG	0.14	(0.34)	0.09	(0.29)	0.17	(0.37)
Risk and time preferences						
Willingness to take risks 0-low, 10-very high	5.33	(2.26)	5.27	(2.37)	5.38	(2.18)
Very impulsive	0.30	(0.46)	0.34	(0.47)	0.28	(0.45)
Very impatient	0.27	(0.44)	0.33	(0.47)	0.22	(0.42)
Observations	986		452		534	

Notes: SOEP v30, 2002–2013, weighted. $^{\natural}$ = Deflated to base year 2007 and in hundreds of Euro.

While about one third of the parents lived in the former GDR in 1989, the descriptive difference between takers and non-takers is considerable: The percentage of students with East German background is two thirds higher in the group of those who claim the benefits and the difference is highly statistically significant.

The same is true for older siblings as a potential source of support in filing the BAföG application: The percentage of claimants in the group of students with older siblings who have already claimed is twice as large as the percentage of those who cannot draw upon older siblings' experiences ($p < 0.01$).

Finally, the percentage of the students rating themselves as very impulsive and impatient is higher in the group of students who turn down the benefits, whereas the willingness to take risk does not differ significantly ($p > 0.1$).

3.6 Non-take-up of BAföG

3.6.1 Estimated rates of non-take-up

Figure 3.2 reveals that about two in five students do not claim BAföG, though eligible; the non-take-up rates range between 36% (NTU_L) and 40% (NTU) on average. Reassuringly, both rates do not differ much so that the impact of potentially misclassified cases should be low.[20]

Moreover, we do not find statistically significant differences in the NTUs (and beta error) over time, which reassures us once more that the non-separability of BAföG and scholarships through 2006 is not an issue.[21]

To shed some more light on the relationships between our main variables, we plot the deflated BAföG amounts from our microsimulation against the parents' deflated last year's monthly net household income (figure 3.3). To account for scale effects in consumption within the household, we use the modified OECD equivalence scale. The simulated funding amounts for eligible students, i.e., students with positive amounts, are depicted in dark grey, the zero funding amounts for students fulfilling only the formal criteria in light grey. As expected, the relationship between both variables is negative with students from more affluent families being eligible for lower or zero

[20] As our sensitivity check in section 3.9.3 shows, relaxing our restrictive assumptions decreases the beta error rate substantially. As these manual modifications do not affect the regression results, we present the conservative results without any manual corrections only. Corrected results are available upon request.

[21] Although we do not find evidence for a time-trend or statistically significant differences through 2006, we include separate year-dummies in all our regressions.

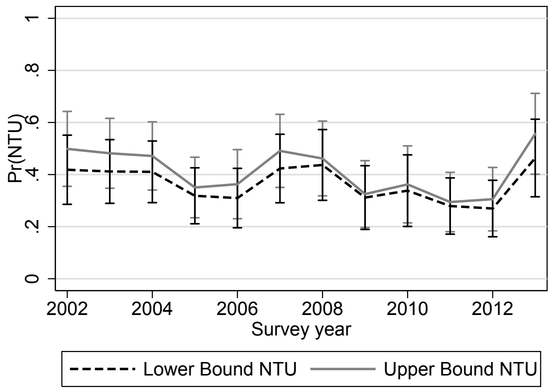


Figure 3.2: The development of the upper and lower bound of the non-take-up rate of BAföG over time

Notes: SOEP v30, 2002–2013, weighted with individual weights, without further controls. The spikes indicate 95% confidence intervals.

funding amounts. At the same time, the variance in BAföG amounts over parents' equivalized income is high as it is not the income used for the BAföG calculation. All in all, our microsimulation model seems to work well in calculating sensible BAföG amounts and yields results comparable to microsimulations from the SOEP-STSM (Steiner and Wrohlich, 2012, p. 130).

Moreover, we investigate which percentage of students is eligible by parents' income and whether eligible students from the lowest tail of the income distribution, where benefits are higher, claim more often than eligible students from higher income families, where benefits are lower (see also Bargain et al. (2012)). Figure 3.4 shows the eligible students' percentage of all formally eligible students, the average benefit amounts of eligible students, and both NTUs up to the 80% percentile of their parents' household equivalized incomes in the previous year (modified OCED-equivalent).

As can be seen from the grey dashed and dotted curves, BAföG is theoretically well targeted to the students from families with low income and/or many children. Accordingly, nearly all students up to the second decile of parents' equivalized income are eligible to positive funding of EUR 400 on average. After the third decile, the curves of the probability to be eligible and the average funding levels slope steeply downward until less than 20% of the

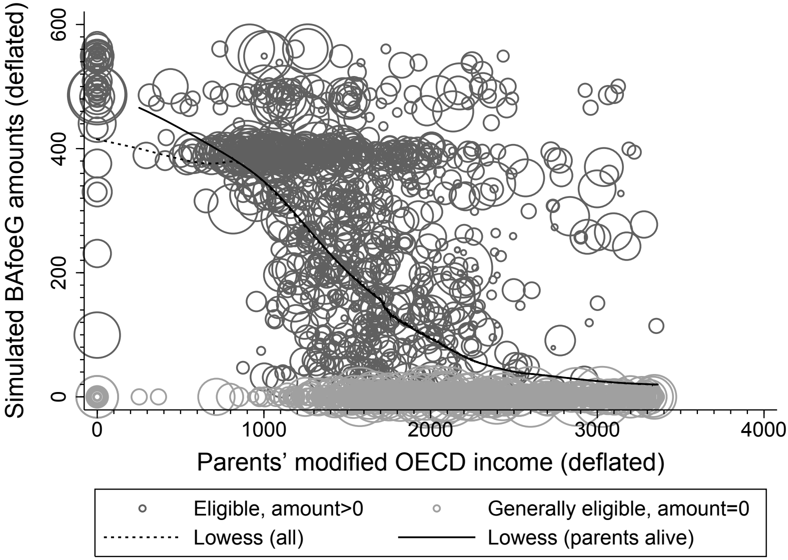


Figure 3.3: Simulated amounts of BAföG benefits over parents' monthly household equivalized income

Notes: SOEP v30, 2002–2013, weighted with individual weights, without further controls. Parents' monthly equivalized household income (modified OECD-scale) is deflated to base year 2007 and presented here if it is below EUR 3728, i.e., below the sum of the mean and one standard deviation of the equivalized household income. The equivalized household income is zero if both parents are deceased but the student is independently funded. The data are weighted so that the relative size of the circles indicate how much weight a respective observation, having been over- or underrepresented in the SOEP, receives. Larger circles indicate that the respective observation receives relatively more weight.

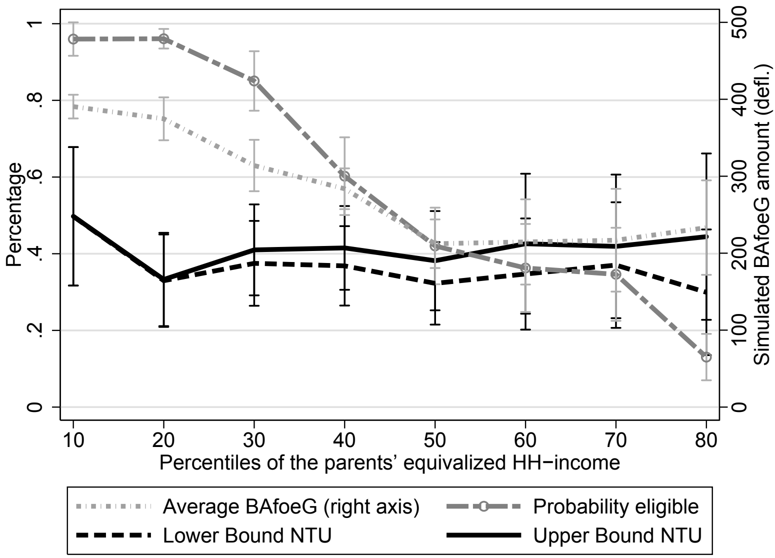


Figure 3.4: Non-take-up rate of BAföG and probability to be eligible by percentiles of the parents' equivalized household income

Notes: SOEP v30, 2002–2013, weighted with individual weights, without further controls and cluster-robust standard errors. Parents' monthly equivalized household income (modified OECD-scale) and the BAföG amount are deflated to base year 2007. Cases whose parental income is not relevant for the BAföG calculation are excluded.

students are eligible to an average amount of EUR 270 in the eighth decile. Factual target efficiency is, however, low: Students from the poorest families are as likely to forgo aid than students from households with higher incomes who are eligible to lower benefits. These results already suggest the limited contribution of parents' income and the level of funding available to explain why a large percentage of the students does not take up BAföG.

3.6.2 Factors of non-take-up

In this section, we want to investigate more closely why students turn down high subsidies. Table 3.2 gives an overview over coefficients and average marginal effects (AME) from our multivariate analyses. We start with discussing the AMEs from the pooled Probit model in column 1 first, and outline later differences with respect to the IV Probit (column 2), the TSLS model (column 3), and the Heckprobit model (column 4).

The average baseline predicted probability of a student not to take up BAföG is about 42%, which is roughly in line with estimates from the literature on the NTU of social assistance in Germany reviewed by Bruckmeier et al. (2013).

For every EUR 100 of benefits available each month, the probability to turn down BAföG decreases by rather modest 4.4 percentage points (13.8%) on average. Accordingly, the elasticity of the level of benefits with respect to the NTU implies that an increase in BAföG by 10% decreases the probability not to take up by 4.6%. To assess the economic significance of increases in the level of the benefits further, we calculate the AME of changing BAföG from the 5th to the 95th percentile, keeping all other variables at their observed values: On average, the probability not to take up BAföG decreases by roughly 20 percentage points from Pr(NTU=1)=0.54 to Pr(NTU=1)=0.33 when BAföG increases from EUR 48 to EUR 500 ($p < 0.05$).

Table 3.2: Different specifications for the predicted probability not to take up BAföG, i.e., $\Pr(NTU = 1|\mathbf{X})$

	Probit		IV Probit		TSLS	Heckprobit	
	(1) Coeff	(2) AME	(3) Coeff	(4) AME	(5) Coeff	(6) Coeff	(7) AME
Simulated BAfoeG amount[♮]	−0.137**	−0.044***	−0.150**	−0.048**	−0.048**	−0.133**	−0.043***
	(0.054)	(0.017)	(0.062)	(0.019)	(0.020)	(0.053)	(0.016)
Age (centered)	−0.003	−0.001	−0.002	−0.001	−0.001	0.017	0.005
	(0.034)	(0.011)	(0.034)	(0.011)	(0.011)	(0.033)	(0.011)
Female	−0.095	−0.030	−0.096	−0.031	−0.037	−0.078	−0.025
	(0.144)	(0.046)	(0.144)	(0.046)	(0.048)	(0.139)	(0.045)
Migration background	−0.108	−0.034	−0.104	−0.033	−0.038	−0.182	−0.058
	(0.209)	(0.066)	(0.210)	(0.066)	(0.072)	(0.203)	(0.064)
Living situation controls							
Student living in urban area	0.607***	0.190***	0.607***	0.190***	0.203***	0.563***	0.180***
	(0.163)	(0.049)	(0.163)	(0.049)	(0.052)	(0.161)	(0.050)
Student living at parents' home	0.838***	0.268***	0.834***	0.267***	0.283***	0.853***	0.277***
	(0.195)	(0.058)	(0.195)	(0.058)	(0.062)	(0.192)	(0.057)
Student lives in East Germany	0.287	0.092	0.293	0.094	0.094	0.308	0.099
	(0.234)	(0.074)	(0.234)	(0.074)	(0.073)	(0.234)	(0.073)
Parent and sibling controls							
Log parental current gross labor income[♮]	−0.031	−0.010	−0.036	−0.012	−0.011	−0.014	−0.005
	(0.056)	(0.018)	(0.057)	(0.018)	(0.018)	(0.055)	(0.018)
Parent(s) have college degree	−0.132	−0.042	−0.137	−0.044	−0.041	−0.113	−0.037
	(0.154)	(0.049)	(0.154)	(0.049)	(0.050)	(0.149)	(0.048)
Parents received social transfers	−0.265	−0.084	−0.260	−0.082	−0.085	−0.269	−0.086
	(0.203)	(0.063)	(0.203)	(0.063)	(0.064)	(0.198)	(0.062)
East German background	−0.458**	−0.148**	−0.456**	−0.147**	−0.159**	−0.523***	−0.170***
	(0.203)	(0.065)	(0.204)	(0.065)	(0.062)	(0.202)	(0.065)
Older sibling claimed BAfoeG	−0.677***	−0.204***	−0.680***	−0.205***	−0.239***	−0.712***	−0.218***
	(0.192)	(0.053)	(0.193)	(0.053)	(0.062)	(0.190)	(0.054)

Table 3.2: Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk and time preferences							
Willingness to take risks (std)	−0.096	−0.031	−0.095	−0.030	−0.033	−0.094	−0.031
	(0.068)	(0.022)	(0.067)	(0.021)	(0.022)	(0.066)	(0.021)
Very impulsive	−0.098	0.032	−0.104	0.030	−0.027	−0.076	0.043
	(0.199)	(0.053)	(0.198)	(0.053)	(0.063)	(0.194)	(0.053)
Very impatient	−0.005	0.068	−0.003	0.069	0.010	−0.060	0.056
	(0.238)	(0.059)	(0.238)	(0.059)	(0.079)	(0.234)	(0.058)
Very impulsive × Very impatient	0.695**		0.694*		0.228*	0.744**	
	(0.353)		(0.354)		(0.116)	(0.345)	
Instruments (1st stage)							
Individual max. BAfoeG amount			0.934***				
			(0.022)				
Independently funded			0.516***				
			(0.155)				
Exclusion restriction (1st stage)							
Vocational training completed						−0.675***	
						(0.231)	
Observations	986		986		986	1041	
Baseline predicted probability	0.417		0.417		0.416	0.444	
corr(u1,u2)=ρ			0.041			−0.748	
Wald test ($\rho = 0$, p-value)			0.526			0.052	
Robust score test (p-value)					0.464		
Overidentification test (p-value)			0.728†		0.353		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, 2002–2013, weighted. All regressions include year controls. ♮ = Deflated to base year 2007 and in hundreds of Euro. † p-value from a J overidentification test on an unweighted, twostep version of the IV Probit without cluster-robust standard errors; estimated with the weakiv package in Stata (Finlay et al., 2013).

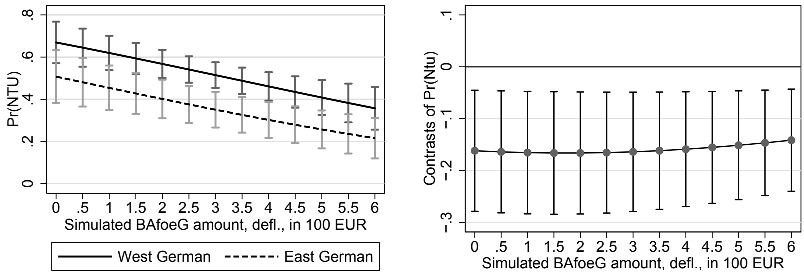
The controls for the students' living situation reveal that students living in urban areas with, presumably, more employment opportunities are about 19 percentage points more likely not to claim BAföG. Those who profit from low living costs because they live at their parents' homes are 27 percentage points more likely not to take up BAföG, whereas living in East Germany does not significantly affect NTU, although the coefficient points to the expected direction.

Investigating our proxies for information constraints, complexity of claiming, and parents' receipt of welfare benefits reveals two things: First, students from families where another social transfer has been claimed in the previous year are less likely to forgo BAföG funding. Yet, the effect is not statistically significantly different from zero. Second, although neither having migration background nor parents' educational and financial situation affect the students' take up decision significantly, having an older sibling who has claimed BAföG before decreases the NTU by 20 percentage points. The latter suggests that support in managing the complex paperwork involved when claiming BAföG is beneficial.

Moreover, there is strong support for our hypothesis that non-take up differs between students socialized in East and those socialized in West Germany. On average, students with an East-German background are about 15 percentage points less likely to reject the money, ceteris paribus. The difference in non-take-up between East and West German background is large and statistically different from zero at $p < 0.05$ over the whole range of the BAföG benefits as displayed in figure 3.5.[22] We closer investigate the robustness of this finding in section 3.7.

With respect to the importance of time-inconsistent preferences, we find a statistically significant interaction of impulsivity and impatience in the expected direction of self-commitment to avoid overspending. In table 3.3, we show the predicted probabilities of NTU for high and low levels of impulsivity and impatience, keeping all other variables at their observed values. The predicted probabilities of students who are high in impatience and low in impulsivity or vice versa do not differ significantly. Impatient students who are very impulsive at the same time are, however, about 23 percentage points more likely to reject the same benefit amount than are impulsive but patient students. This difference is highly statistically significant. We find a symmetrical effect of about 20 percentage points for impatient students when we vary the level of impulsivity. The large double difference of about 23 percentage points (which represents the size of the interaction effect in terms

[22] The gap is also robust to introducing an interaction between East German background and parents' incomes to our model, although this results in a high degree of multicollinearity.



(a) Predicted probabilities (b) Contrasts of predicted probabilities

Figure 3.5: Impact of socialization on non-take-up of BAföG by simulated benefits and by whether parents lived in East or West Germany in 1989

Notes: SOEP v30, 2002–2013, weighted with individual weights. The spikes indicate 95% confidence intervals. Predicted probabilities were calculated from the Probit regression in table 3.2, column 1. All other variables were held at their observed values.

of AMEs) is also statistically significantly different from zero ($p < 0.05$) and in line with the sign and significance we find for the interaction effect in terms of our Probit coefficients. To ensure that the effect is meaningful over the whole range of BAföG amounts, we calculated contrasts for every EUR 50 of the BAföG amount as shown in figure 3.6a. The difference is stable and statistically significantly different from zero over the whole range of possible funding amounts (figure 3.6b). All in all, our results yield strong evidence for the hypothesis that students with self-control problems restrict their future funding sources as to avoid overspending. As expected, willingness to take risks is not associated with non-take-up.

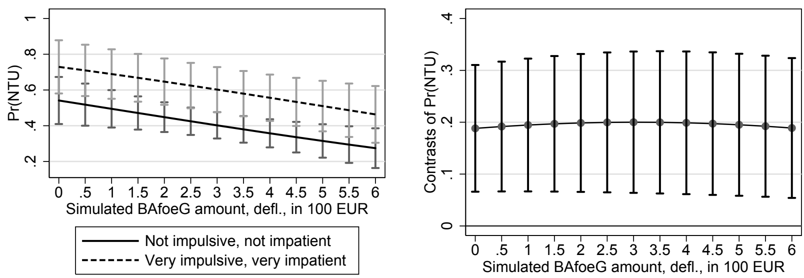
The second and third columns in table 3.2 present the results from running instrumental variable regressions for the Probit (col. 2) and the linear probability model case (col. 3), using the individual maximum benefits amount and an indicator for whether the student is independently funded as instruments. As indicated by the Wald test of exogeneity and Wooldridge (1995)'s robust score test, we do not find evidence for potential endogeneity of the benefits amount, neither in the non-linear nor in the linear model. In line with this and against the background that our correlation in the errors (u_1, u_2) in the IV Probit is very low, our results are, by and large, unaffected by whether we account for the potential endogeneity of the benefits amount or not. As IV Probit and TSLS are also very similar, the somewhat stronger distributional

Table 3.3: Predicted probabilities for non-take-up of BAföG by different levels of the students' impulsivity and impatience

		Very impulsive		
		No	Yes	Difference
Very impatient	No	0.397***	0.366***	-0.032
		(0.037)	(0.058)	(0.064)
	Yes	0.396***	0.594***	0.199**
		(0.078)	(0.064)	(0.095)
	Difference	-0.002	0.229***	0.230**
		(0.077)	(0.083)	(0.116)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Unconditional, cluster-robust standard errors in parentheses.

Notes: SOEP v30, 2002–2013, weighted. Predicted probabilities of the Probit in table 3.2, col. 1. All other variables were kept at their observed values.



(a) Predicted probabilities (b) Contrasts of predicted probabilities

Figure 3.6: Impact of impulsiveness and impatience on non-take-up of BAföG by the simulated benefit amount

Notes: SOEP v30, 2002–2013, weighted with individual weights. The spikes indicate 95% confidence intervals. Predicted probabilities were calculated from the Probit regression in table 3.2, column 1. All other variables were held at their observed values.

assumptions of the IV Probit do not harm our results. Reassuringly, the first stage coefficients and p-values reported at the bottom of the table indicate that both instruments are very strong—as does a Shea's Adjusted Partial R-squared of .80 from the first stage of the TSLS.[23] Because our model is overidentified, we can conditionally test the exogeneity assumption with an overidentification test. As reported at the bottom of table 3.2, we cannot reject the hypothesis that the additional instrument is exogenous.

We check whether our specification in column 1 is affected by self-selection in column 4, where we report results from our Heckman-type Probit sample selection model. Only few students dropped out of our sample because they had too much assets or income. Nevertheless, our hypothesis that the errors of regression and selection equation are not correlated is rejected at $p = 0.05$. The correlation of the errors (u_1, u_2) is moreover negative as is the highly statistically significant exclusion restriction, suggesting that students who completed vocational training before studying have a lower probability to remain in our sample of eligible. Although we find evidence that sample selection is an issue, the resulting AMEs, especially for the benefits level, are very similar to those from the straightforward Probit model, presumably because the number of selected cases is low: The predicted probability to turn down BAföG slightly increases to 44%, and the elasticity of the average non-take-up probability with respect to a 10% increase in the benefits slightly reduces to 4.1%. The impact of East German background, siblings' claiming experience, and debt aversion is somewhat more pronounced. All other conclusions we have drawn from the Probit model (column 1) remain valid.

Taken together, our results suggest that most students stay roughly within the thresholds used for assessment of BAföG eligibility and family insurance so that we find no evidence for endogeneity of the benefit amount if we restrict our sample to students eligible for funding after own incomes are deducted. Nevertheless, some students are likely to earn so much that they lose their complete eligibility and select themselves out of the sample. This sample-selection should be accounted for, so that the Heckprobit model results in our preferred specification.

[23] It is not straightforward how to test for weak instruments in pooled non-linear models with cluster-robust standard errors and weighted data because there is no clear cut-off for non-linear models to guide us when to reject the hypothesis of weak instruments. Yet, a Kleibergen-Paap Wald F statistic of 1376.11 from our weighted TSLS with cluster-robust standard errors greatly exceeds the Stock and Yogo (2005) critical values of F=19.93 for a relative bias of 10% and provides additional evidence that the instruments are relevant.

We run separate analyses to investigate the effect of the duration of benefits as including this variable reduces our sample again.[24] As expected, the relationship between a high number of semesters and non-take-up is positive, but slightly decreasing as we consider only students in the eligible semester range (table 3.4): The more advanced the student is in her studies, the higher the probability that she does not take up the benefits because the period in which the claiming costs pay off is shorter.

3.7 Robustness checks

3.7.1 Different welfare preferences

The stable difference in NTU between students socialized in the East and in the West might be either a masked difference in scale effects or the "welfare trap", given that East and West Germans differ significantly in claiming other social benefits. Therefore, we add an interaction between our East German background variable and the social benefit dummy to our preferred model, the Heckprobit specification, and report results from the Probit as a benchmark. Table A3.13 in the appendix displays the full results. We again report predicted probabilities with their respective differences in table 3.5.

Table 3.5 highlights that the AME of the interaction effect equals -0.15 and is not statistically significant from zero (as are the coefficients of the interactions in table A3.13). Accordingly, having drawn upon other social transfers does not affect families with East and West German background differently.

3.7.2 Parents' financial support

As discussed earlier, the official BAföG calculation uses parents second-last year's incomes, unless students request to use parents' last year's or current incomes. Our microsimulation model is, therefore, based on the assumption that students request an update to more recent incomes if these are lower. If parents' income grows very fast and if parents use the surplus to support

24 The microsimulation accounts for the fact that only students in a certain range of semesters are eligible to receive BAföG. We keep observations with missing information on the year of enrollment in higher education in our sample used for the previous analyses if students report to claim BAföG, assuming that they should accordingly still fall into the eligible range of semesters. Inclusion of these observations does not affect our results.

Table 3.4: Duration of receipt and non-take-up probability Pr($NTU = 1|\mathbf{X}$)

	(1) Probit		(2) Heckprobit	
	Coeff	AME	Coeff	AME
Simulated BAfoeG amount$^\natural$	−0.151***	−0.049***	−0.145**	−0.047***
	(0.058)	(0.018)	(0.058)	(0.018)
Female	−0.086	−0.028	−0.076	−0.025
	(0.147)	(0.048)	(0.142)	(0.046)
Migration background	−0.084	−0.027	−0.147	−0.047
	(0.217)	(0.069)	(0.209)	(0.067)
Academic year	0.376***	0.044***	0.342***	0.042***
	(0.116)	(0.016)	(0.112)	(0.015)
Academic year^2	−0.047***		−0.042***	
	(0.016)		(0.016)	
Living situation controls				
Student living in urban area	0.574***	0.183***	0.522***	0.169***
	(0.167)	(0.052)	(0.163)	(0.052)
Student living at parents' home	0.856***	0.277***	0.830***	0.273***
	(0.188)	(0.056)	(0.184)	(0.056)
Student lives in East Germany	0.282	0.090	0.304	0.097
	(0.243)	(0.077)	(0.240)	(0.075)
Parent and sibling controls				
Log parental current gross labor income$^\natural$	−0.041	−0.013	−0.028	−0.009
	(0.056)	(0.018)	(0.054)	(0.018)
Parent(s) have college degree	−0.133	−0.043	−0.105	−0.034
	(0.163)	(0.053)	(0.155)	(0.050)
Parents received social transfers	−0.264	−0.085	−0.281	−0.091
	(0.213)	(0.067)	(0.207)	(0.066)
East German background	−0.465**	−0.151**	−0.528***	−0.174**
	(0.209)	(0.068)	(0.205)	(0.068)
Older sibling claimed BAfoeG	−0.663***	−0.205***	−0.700***	−0.220***
	(0.201)	(0.057)	(0.199)	(0.058)
Time-inconsistent preferences				
Willingness to take risks (std)	−0.052	−0.017	−0.048	−0.016
	(0.070)	(0.022)	(0.068)	(0.022)
Very impulsive	−0.066	0.027	−0.051	0.036
	(0.201)	(0.055)	(0.196)	(0.054)
Very impatient	0.031	0.063	−0.021	0.052
	(0.245)	(0.061)	(0.241)	(0.060)
Very impulsive × Very impatient	0.530		0.590*	
	(0.356)		(0.348)	
Exclusion restriction (1st stage)				
Vocational training completed			−0.903***	
			(0.203)	
Year controls	✓		✓	
Observations	944		998	
Baseline predicted probability	0.442		0.470	
corr(u1,u2)=ρ			−0.844	
Wald test ($\rho = 0$, p-value)			0.057	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, 2002–2013, weighted. $^\natural$ = Deflated to base year 2007 and in hundreds of Euro.

Table 3.5: Predicted probabilities for non-take-up of BAföG by the students' East German background and whether parents received other social transfers last year

		Other social transfer		
		No	Yes	Difference
East German Background	No	0.501***	0.482***	-0.020
		(0.042)	(0.090)	(0.089)
	Yes	0.366***	0.196***	-0.170**
		(0.057)	(0.074)	(0.081)
	Difference	-0.135*	-0.286**	-0.150
		(0.069)	(0.117)	(0.120)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Unconditional, cluster-robust standard errors in parentheses.

Notes: SOEP v30, 2002–2013, weighted. Predicted probabilities of the Heckprobit in table A3.13, col. 2. All other variables were kept at their observed values.

their children financially, we might overestimate the students' needs and, accordingly, the importance of the level of BAföG benefits. This biases our results only if the factors of income growth are not controlled for by the socio-economic covariates in our model, and if the income growth is related to an disproportional increase of the direct transfers to the offspring.

We add an indicator for whether parents supported the student financially to columns 1 and 2 of table 3.6, which results in the loss of one observation due to item non-response. The indicator is not statistically significantly different from zero and does not affect the other coefficients much.

3.7.3 Different simulation quality

To rule out the possibility that our evidence of non-take-up is simply resulting from poorer data quality for some cases, we construct indicators for whether parents' income is imputed by the SOEP and whether students, parents, or spouses/partners round their gross income to EUR 100. As shown in table 3.7, these indicators are not statistically significantly different from zero and provide no evidence that different simulation quality introduces bias to our estimates.

Table 3.8 investigates whether the estimates differ when we restrict our sample to those for whom we can differentiate between merit-based scholar-

Table 3.6: Parents' financial support does not impact on non-take-up

	(1) Probit		(2) Heckprobit	
	Coeff	AME	Coeff	AME
Simulated BAfoeG amount♮	−0.128**	−0.041**	−0.127**	−0.041***
	(0.052)	(0.016)	(0.051)	(0.016)
Age (centered)	−0.006	−0.002	0.014	0.005
	(0.033)	(0.011)	(0.033)	(0.011)
Female	−0.087	−0.028	−0.072	−0.023
	(0.144)	(0.046)	(0.140)	(0.045)
Migration background	−0.094	−0.030	−0.170	−0.054
	(0.211)	(0.066)	(0.204)	(0.064)
Living situation controls				
Student living in urban area	0.620***	0.194***	0.575***	0.183***
	(0.164)	(0.049)	(0.161)	(0.049)
Student living at parents' home	0.908***	0.288***	0.909***	0.293***
	(0.206)	(0.059)	(0.205)	(0.060)
Student lives in East Germany	0.279	0.089	0.301	0.096
	(0.235)	(0.074)	(0.234)	(0.074)
Parent and sibling controls				
Log parental current gross labor income♮	−0.037	−0.012	−0.019	−0.006
	(0.056)	(0.018)	(0.055)	(0.018)
Parent(s) have college degree	−0.128	−0.041	−0.109	−0.035
	(0.154)	(0.049)	(0.148)	(0.048)
Parents received social transfers	−0.268	−0.085	−0.273	−0.087
	(0.205)	(0.063)	(0.200)	(0.063)
East German background	−0.458**	−0.147**	−0.521***	−0.170***
	(0.204)	(0.065)	(0.201)	(0.065)
Older sibling claimed BAfoeG	−0.693***	−0.208***	−0.724***	−0.221***
	(0.191)	(0.053)	(0.190)	(0.053)
Parents' financial support last year	0.161	0.051	0.127	0.041
	(0.147)	(0.046)	(0.146)	(0.046)
Time-inconsistent preferences				
Willingness to take risks (std)	−0.103	−0.033	−0.100	−0.032
	(0.067)	(0.021)	(0.065)	(0.021)
Very impulsive	−0.099	0.032	−0.077	0.043
	(0.198)	(0.053)	(0.194)	(0.053)
Very impatient	−0.001	0.070	−0.056	0.058
	(0.237)	(0.059)	(0.233)	(0.058)
Very impulsive × Very impatient	0.693**		0.743**	
	(0.353)		(0.344)	
Exclusion restriction (1st stage)				
Vocational training completed			−0.708***	
			(0.235)	
Year controls	✓		✓	
Observations	985		1040	
Baseline predicted probability	0.416		0.443	
corr(u1,u2)=ρ			−0.753	
Wald test ($\rho = 0$, p-value)			0.039	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, weighted. ♮ = Deflated to base year 2007 and in hundreds of Euro.

ship receipt and BAföG receipt, i.e., we limit the sample to those surveyed after 2006. Our point estimates are, overall, similar to those from the full sample. We find, however, no evidence for a significant sample selection bias—most probably because the number of cases with self-selection is too low.

3.7.4 Further robustness checks

We want to mention briefly that our results are also robust to several other robustness checks (results available upon request):

First, until August 2015, students who were only preliminarily accepted for their consecutive studies faced problems receiving BAföG without interruptions, e.g., when applying for a Master program before having completed the Bachelor's thesis. The number of students in our sample who are enrolled in consecutive programs is, however, very low. Excluding these cases does not affect our results.

Second, the introduction and the abolition of tuition fees of up to EUR 500 per semester at several German universities in some federal states fall into our observation window. BAföG recipients were, generally, also obligated to pay the fees and their parents were expected to increase their financial support accordingly if possible. Evidence on whether the introduction of the fees had an effect is mixed (Hübner, 2012; Bruckmeier and Wigger, 2014). We construct an indicator based on the students' place of living in a certain year and merge information from federal amendments indicating which federal state introduced tuition fees in which year. The indicator is not statistically significantly different from zero and its inclusion does not affect our results.

Third, we investigate different specifications of our model. Adding further variables to our models in table 3.2 (student married, age squared, parents' relationship, student receives parental financial support, student has siblings, parents had debts last year, Big Five personality traits, desired age of economic independence as reported at age 17) neither increases model fit nor provides any indication of potential omitted variable bias, so that we report the most parsimonious models only. Moreover, using a broader measure for the parents' income, such as the parents' household net income, does not affect the results. Last, we find no indication of enough non-linearity in the data to justify higher order polynomials of the BAföG amount.

Table 3.7: Missing data does not impact on non-take-up (full sample)

	(1) Probit		(2) Heckprobit	
	Coeff	AME	Coeff	AME
Simulated BAfoeG amount♮	−0.133**	−0.042**	−0.129**	−0.042**
	(0.054)	(0.017)	(0.053)	(0.016)
Age (centered)	−0.005	−0.002	0.016	0.005
	(0.034)	(0.011)	(0.033)	(0.011)
Female	−0.095	−0.030	−0.081	−0.026
	(0.145)	(0.046)	(0.140)	(0.045)
Migration background	−0.113	−0.036	−0.189	−0.060
	(0.209)	(0.066)	(0.203)	(0.064)
Living situation controls				
Student living in urban area	0.635***	0.198***	0.580***	0.184***
	(0.161)	(0.048)	(0.159)	(0.049)
Student living at parents' home	0.831***	0.266***	0.841***	0.273***
	(0.195)	(0.058)	(0.192)	(0.058)
Student lives in East Germany	0.313	0.100	0.330	0.105
	(0.234)	(0.074)	(0.232)	(0.073)
Parent and sibling controls				
Log parental current gross labor income♮	−0.045	−0.014	−0.028	−0.009
	(0.056)	(0.018)	(0.055)	(0.018)
Parents have college degree	−0.130	−0.042	−0.110	−0.036
	(0.154)	(0.049)	(0.148)	(0.048)
Parents received social transfers	−0.273	−0.086	−0.273	−0.087
	(0.203)	(0.063)	(0.197)	(0.062)
East German background	−0.471**	−0.151**	−0.537***	−0.175***
	(0.204)	(0.065)	(0.201)	(0.065)
Older sibling claimed BAfoeG	−0.689***	−0.207***	−0.727***	−0.222***
	(0.195)	(0.054)	(0.192)	(0.054)
Data-quality indicators				
Parents' income imputed	0.070	0.023	0.060	0.019
	(0.059)	(0.019)	(0.058)	(0.019)
Gross income rounded	0.088	0.028	0.110	0.035
	(0.128)	(0.041)	(0.125)	(0.040)
Risk and time preferences				
Willingness to take risks (std)	−0.097	−0.031	−0.095	−0.031
	(0.068)	(0.022)	(0.066)	(0.021)
Very impulsive	−0.117	0.028	−0.092	0.039
	(0.200)	(0.053)	(0.195)	(0.052)
Very impatient	−0.010	0.070	−0.067	0.056
	(0.241)	(0.059)	(0.235)	(0.058)
Very impulsive × Very impatient	0.720**		0.764**	
	(0.355)		(0.345)	
Exclusion restriction (1st stage)				
Vocational training completed			−0.723***	
			(0.227)	
Observations	986		1041	
Baseline predicted probability	0.417		0.444	
corr(u1,u2)=ρ			−0.789	
Wald test ($\rho = 0$, p-value)			0.042	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, weighted. ♮ = Deflated to base year 2007 and in hundreds of Euro. All regressions include year controls.

Table 3.8: Simulation quality does not impact on non-take-up (sample after 2006)

	(1) Probit		(2) Heckprobit	
	Coeff	AME	Coeff	AME
Simulated BAfoeG amount♮	−0.115*	−0.037*	−0.116*	−0.037*
	(0.067)	(0.021)	(0.067)	(0.021)
Age (centered)	0.007	0.002	0.020	0.006
	(0.045)	(0.014)	(0.048)	(0.015)
Female	−0.019	−0.006	0.006	0.002
	(0.180)	(0.057)	(0.179)	(0.057)
Migration background	−0.095	−0.030	−0.139	−0.044
	(0.253)	(0.079)	(0.245)	(0.076)
Living situation controls				
Student living in urban area	0.656***	0.200***	0.627***	0.193***
	(0.195)	(0.056)	(0.205)	(0.059)
Student living at parents' home	0.799***	0.254***	0.834***	0.267***
	(0.248)	(0.073)	(0.251)	(0.075)
Student lives in East Germany	0.576**	0.183**	0.595**	0.187**
	(0.278)	(0.085)	(0.281)	(0.084)
Parent and sibling controls				
Log parental current gross labor income♮	−0.056	−0.018	−0.045	−0.014
	(0.080)	(0.025)	(0.081)	(0.026)
Parents have college degree	−0.070	−0.022	−0.066	−0.021
	(0.193)	(0.061)	(0.190)	(0.060)
Parents received social transfers	−0.400	−0.123	−0.393	−0.122
	(0.267)	(0.079)	(0.265)	(0.079)
East German background	−0.560**	−0.175**	−0.597**	−0.188***
	(0.235)	(0.072)	(0.237)	(0.073)
Older sibling claimed BAfoeG	−0.646***	−0.192***	−0.683***	−0.204***
	(0.228)	(0.062)	(0.228)	(0.063)
Risk and time preferences				
Willingness to take risks (std)	−0.126	−0.040	−0.127	−0.040
	(0.079)	(0.025)	(0.078)	(0.025)
Very impulsive	−0.220	−0.002	−0.206	0.005
	(0.236)	(0.062)	(0.239)	(0.064)
Very impatient	0.058	0.099	0.022	0.093
	(0.277)	(0.073)	(0.274)	(0.072)
Very impulsive × Very impatient	0.784*		0.829**	
	(0.425)		(0.421)	
Exclusion restriction (1st stage)				
Vocational training completed			−0.848***	
			(0.308)	
Observations	625		659	
Baseline predicted probability	0.401		0.422	
corr(u1,u2)=ρ			−0.460	
Wald test ($\rho = 0$, p-value)			0.533	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, 2007–2013, weighted. ♮ = Deflated to base year 2007 and in hundreds of Euro. All regressions include year controls.

3.8 Discussion

This chapter investigates which percentage of eligible students do not take up the German student financial aid, BAföG, and provides insights into the explanatory factors of non-take-up. We explicitly account for endogeneity of the level of benefits and students selecting themselves out of the group of eligible.

Although the combination of a grant and zero interest loan is very lucrative and classical economics would expect students to claim the aid amounts, about two fifths of the students forgo funding. Students are more likely to claim the benefits if the expected duration of funding is high. Moreover, increasing the level of benefits by 10% reduces the probability of non-take-up by about 4.1% on average when sample selection is taken into account. The probability of non-take-up is, therefore, relatively inelastic with respect to the level of benefits, though our estimate is about one third higher than those found for non-take-up of social assistance in Germany (Anderson and Meyer, 1997; Riphahn, 2001). Our evidence of the existence of BAföG non-take-up and its rather low benefit-level elasticity provide a novel explanation why increasing the level of student financial aid cannot raise students' university enrollment substantially (Baumgartner and Steiner, 2005, 2006; Steiner and Wrohlich, 2012).

We test hypotheses on various factors related to non-take-up behavior. We find that students socialized in the former socialist East, where people still have stronger preferences for high levels of social security and equality, are considerably less likely to forgo the benefits, irrespective of whether parents claimed other welfare benefits in the previous year. At the same time, students with siblings who already claimed the benefits and are, thus, acquainted with the formalities of claiming are more than 20 percentage points more likely to take up BAföG. Debt aversion, to the contrary, is strongly associated with higher probabilities of non-take-up.

Like most other studies investigating non-take-up of social benefits, we have to rely on survey data to draw upon information of both eligible and ineligible students and to be able to shed light on the reasons for non-take up. The use of survey data is, however, associated with well-known limitations such as measurement error or small sample sizes for specific subgroups.[25]

Furthermore, as survey data usually lacks direct measures of the reasons to reject social benefits, we have to base our analyses on proxy variables that generally yield conflicting expectations about the theoretical direction of

[25] See Hernanz et al. (2004) for an extensive overview over (dis-)advantages of various data sources for the analysis of non-take-up.

the effects (Becker and Hauser, 2003, p. 149f) or do not allow to disentangle competing explanations. We carefully account for potential endogeneity arising from students' endogenous choices of their incomes and do, therefore, implicitly incorporate unobserved differences in, e.g., abilities or motivation. Nevertheless, we cannot rule out that some degree of omitted variable bias remains. More specifically, further research is needed to assess whether other behavioral economics explanations for students' non-take-up of BAföG do also matter, for example procrastination, mental accounting, or framing effects (see Boatman et al. (2014) for an overview over the last two channels).

Due to our rather small sample size and the low within- and between-variation, we restrict our analyses to pooled cross-sections. To the best of our knowledge, sufficiently rich data sets allowing to account for unobserved heterogeneity between students in a panel-design are not yet available. Once appropriate data are available, rerunning our analyses in a longitudinal design is an interesting avenue for future research. In case this data included repeated measures of the real incomes and assets or allowed to merge external register data on income, future studies should also account for measurement error as done by, e.g., Hernandez and Pudney (2007).

Up to now, we can, nevertheless, conclude that a significant share of students does not claim the student financial aid available. Non-take-up is potentially detrimental to intergenerational educational mobility if these students prolong their time to degree, graduate with worse grades, or fail to graduate completely. As previous studies suggest (Triventi, 2014, e.g.), dropping out without a degree is often a consequence of financial hardship or long working hours.

Our results suggest that take-up is not easily increased by simply raising the level of benefits. Against the background that we find strong evidence for debt aversion resulting from students' fear to spend the money they intended to save, a policy implication would be to provide only the grant component of BAföG as a default, instead of automatically embedding the loan into the BAföG scheme.

Furthermore, filing applications online, which will be possible as of autumn 2016, provides several starting points to facilitate the administrative processes and to decrease the opportunity costs of claiming BAföG. For example, information from students' previous applications or from parents' electronic income tax declaration could be prepopulated.

Finally, we are convinced that simplifying the overly complicated application forms for BAföG would not only cut red tape, but also decrease the number of students who are put off claiming and at risk of financial hardship.

3.9 Appendix

These appendices outline briefly the complex calculation of the BAföG benefits, the rough set-up of our microsimulation model and the sensitivity of the NTU and beta error to necessary assumptions on our data. To reduce the complexity of our descriptions, the following sections refer to the BAföG calculation for a single student. If the student is married or lives in a registered partnership, the calculation is similar, but based on the spouse's or partner's income and living situation instead of on the parents'.

3.9.1 BAföG calculation

To illustrate the BAföG calculation, we draw upon a simple example, representing the standard case (see table A3.9). The example details the BAföG calculation for a single 24-year-old student in 2015 who is not living at her parents' home. The student has neither own earnings, nor assets. Her brother is in vocational training and earns EUR 6,000 per year. Their parents are married; the mother is not employed, but the father earned EUR 38,000 in 2013 and EUR 44,000 in 2015 as an employee in a close-by company. The father's travel distance each day amounts to 5 km for a single journey.

First, we identify the income considered relevant for the calculation of BAföG. The income relevant for BAföG is generally defined as the sum of all positive earnings according to § 2 sect. 1 and 2 of the Income Tax Act: incomes from agriculture and forestry, income from industrial or commercial activities, income from self-employment, employment income, income from investment of capital, rental income, and other income such as life annuities or income from private sales business. Further income as of § 21 sect. 2a and 3 BAföG (earnings taxable outside Germany) must be added; public-sponsored scholarships of up to EUR 300 (e.g., Deutschlandstipendium) are exempt from these deductions. In our example, the starting point is the sum of parents' gross incomes from employment in 2013, i.e., EUR 38,000.

To calculate the parents' incomes relevant for BAföG (§ 21 and § 24 BAföG), the mother's and father's gross positive earnings are reduced by a lump sum for income-related expenses[26], payed taxes such as income tax, church tax, solidarity surcharge, old age percentage reductions (§ 21 sect. 1 BAföG), and by flat-rate social security benefits (§ 21 sect. 2 BAföG). Some forms of grant-aided privately funded pension schemes are also be subtracted.

[26] In case the actual income-related expenses exceed the general lump sum amounts of currently EUR 1,000, the full amount of income-related expenses can be deduced. The same holds for the student.

Table A3.9: Exemplary BAföG calculation

Calculation of the parents' BAföG-relevant income		
Gross income from employment	EUR	38,000.00
./. Income-related tax deductions (lump sum)	EUR	1,000.00
./. Allowances for social insurance payments	EUR	7,881.00
./. Income tax (including church tax and solidarity surcharge)	EUR	3,214.72
= Parents' income relevant for BAföG	EUR	25,904.28
Monthly parents' income relevant for BAföG	EUR	2,158.69
Calculation of the sibling's BAföG-relevant income		
Vocational training pay (monthly)	EUR	500.00
./. Lump sum allowance (see Tz 21.1.32 BAföGVwV)	EUR	140.00
Sibling's income relevant for BAföG	EUR	360.00
Calculating parents' income relevant for BAföG including allowances		
Monthly parents' income relevant for BAföG	EUR	2,158.69
./. Basic allowance for the parents	EUR	1,605.00
./. Basic allowance for the sibling reduced by sibling's income	EUR	125.00
= Parents' income relevant for BAföG reduced by basic allowances	EUR	428.69
./. Additional allowance for parents (50 %) and sibling (5 %)	EUR	235.78
Parents' income relevant for BAföG including allowances	EUR	192.91
Calculating the BAföG amount		
Basic needs	EUR	373.00
+ Rent subsidy	EUR	224.00
= Sum of needs	EUR	597.00
./. Parents' income relevant for BAföG including allowances	EUR	192.91
= BAföG amount	EUR	404.09
BAföG amount (rounded)	EUR	404.00

Notes: Exemplary BAföG calculation for the standard case of a single, childless 24-year-old student, living not at her parents' place and having neither own earnings, nor assets.

In our example, the father's commuting expenses do not exceed the EUR 1,000 lump sum for income-related expenses, so that we reduce his gross income by EUR 1,000 only. We calculate allowances for social insurance payments, church tax, and the solidarity surcharge. After deducting all these components and dividing the sum by 12, we arrive at the parents' monthly "income relevant for BAföG", here equal to EUR 2,158.69. Note that the "income relevant for BAföG" is neither gross nor net income but a special measure used only for the BAföG calculation.

Only the parents' and student's incomes are considered in the calculation of the income relevant for BAföG. Nevertheless, incomes of step parents, children, and other dependents of the parents reduce parents' allowances for children who are not theoretically eligible for BAföG (§ 23 sect. 3 BAföG), see table A3.10. In our example, the student's brother is in vocational training and, therefore, ineligible to claim BAföG. Nevertheless, the parents can protect up to EUR 485 of their incomes to support their son financially. His training pay reduces the parents' maximum allowance, though: The family is allowed to protect a lump sum of monthly EUR 140 from the son's vocational training pay. The remaining EUR 360 are, however, considered as income relevant for BAföG and, therefore, deducted from the maximum parental allowance of EUR 485. All in all, parents can, thus, only deduct an allowance of EUR 125 from their income relevant for BAföG.

Parents can further protect monetary amounts from being means-tested, depending on their living situation. In the case considered here, both parents are married and cohabiting, so that they can protect another EUR 1,605 for their own use. After deducting both the allowance for parents' own use and for their son, the parents' income relevant for BAföG reduced by basic allowances amounts to EUR 428.69. From this amount, parents are, again, granted an additional allowance equal to half of the income relevant for BAföG plus another 5% for each dependent not theoretically eligible for BAföG. As the student's brother is ineligible for BAföG, the parents in our example are granted EUR 235.78 as an additional allowance for themselves and the brother.

Parents are expected to be able to use the remaining amount of EUR 192.91 to support their complete offspring financially. Therefore, the remaining parents' income relevant for BAföG including allowances is divided by the number of dependents formally eligible to receive BAföG. The parents in our example are expected to use the full amount of EUR 192.91 to support their daughter. If the brother had been eligible, the monetary amount of expected support would have dropped to EUR 96.46 as the parents' applicable income would have been divided by two.

Table A3.10: Basic allowances of incomes and assets between 2002 and 2015

Basic allowances of... (in EUR)		2010–2015	2008–2010	2002–2008
...parents' or spouse's / partner's income				
	cohabiting and married or in a registered same-sex partnership	1,605	1,555	1,440
	single parent, not cohabiting	1,070	1,040	520
	additional allowance for step-parent	535	520	480
	additional allowance for children and other dependents ineligible for BAföG	485	470	435
...own income				
	earned income	255	255	112-215
	married or in a registered same-sex partnership and spouse not eligible for BAföG	535	520	480
	with children / dependents ineligible for BAföG (each)	485	470	435
	orphan's pension	125	120	112
	case of hardship	205	205	205
...own assets (per year)				
	for the claimant	5,200	5,200	5,200
	for the spouse / partner	1,800	1,800	1,800
	for each child	1,800	1,800	1,800

Notes: Amounts are in euro and per month (if not indicated otherwise).
Source: Own table based on Rothe and Blanke (2015).

While parents are allowed to protect their full assets—except the interest accruing from it which is part of the sum of positive earnings—, both the student's earnings *and* assets are subject to the means test.

The student's maximum earnings without deductions are calculated as follows (§ 21 and § 23 BAföG): Starting point of the calculation is the student's gross income for the respective year BAföG is claimed for. From this, EUR 1,000 of income-related expenses are subtracted as a lump sum, unless higher factual expenses can be proven. Then, a certain percentage is deduced as a flat-rate amount. The percentage depends on whether the student is compulsorily insured as a student or as an employee in the retirement insurance and on the type of employment. The default is compulsory insurance as a student or as a student working in a job with compulsory insurance, resulting in a flat-rate percentage of 21.3%. Furthermore, to calculate monthly amounts, the remaining amount is divided by 12 months. Last, the respective exempt amounts, depending on the student's living situation (e.g., EUR 255, see table A3.10), are deducted. The maximum gross income to be earned without deductions is, therefore, EUR 4,884 a year or EUR 407 a month if the student is in a minor employment—other than that, the student loses his or her family insurance. With respect to own assets, students are expected to use every euro exceeding a cut-off of EUR 5,200 for their education. As the student in our example has neither own income nor assets, we can skip the means test of own incomes and assets.

To calculate the student's respective BAföG amount, we have to calculate the sum of needs first. The basic need amount equals EUR 373 and can be supplemented by additional amounts, depending on the student's living situation and age, see table A3.11. In our example, the student has her own flat and is, therefore, also eligible to a rent subsidy of EUR 224. Because she is childless and under age 25, she is still insured in her parents' non-contributory dependents' co-insurance and does not qualify for other additional amounts as of table A3.11. From the student's level of needs, we deduct the parents' and student's incomes relevant for BAföG including allowances and the student's assets above EUR 5,200. The resulting amount is the level of monthly benefits to be cashed. For the student in our example, we have to deduct only the parents' income measure and arrive at a rounded BAföG amount of EUR 404.

3.9.2 The microsimulation model

This section explains the basic features of our microsimulation model and details the most important assumptions made on the data.

Table A3.11: Level of needs 2002–2015

Level of needs		2010–2015	2008–2010	2002–2008
Basic need				
	Students in higher education	373	366	333
Additional amounts				
	to cover living expenses if living at home	49	48	44
	to cover living expenses if not living at home	224	146	133
	health insurance	62	50/54	47
	care insurance	11	9/10	8
	first child (below age 10)	113	113	–
	further children (below age 10)	85	85	–

Notes: Amounts are in euro and per month.
Source: Own table based on Rothe and Blanke (2015).

Our microsimulation model takes three main steps to set up the analytic sample used for our analyses: First, we isolate the formally eligible students and their siblings. Second, we prepare our data for the means test. More specifically, we set up an income tax model, roughly following Schwarze (1995), to calculate the incomes relevant for BAföG from the respective gross income amounts. Third, we perform the means test by calculating the students' needs, subtracting the BAföG-relevant incomes, and accounting for the relevant allowances. The third step entails the procedure detailed in section 3.9.1.

As described previously, we determine the formal BAföG eligibility of all students in the SOEP v30, 2002–2013, following § 2 et seq. BAföG. In other words, we have to assess whether students are formally eligible to participate in the means test. While we can easily assess whether students meet the age requirement and are enrolled at an eligible higher education institution,[27] we have to impose assumptions on the maximum funding period. Students can receive funding for their first degree and during the respective average period of studies (*Regelstudienzeit*). The average period of studies varies with the desired degree, the subject of studies, and the type of higher education institution. Lacking full information to construct individual-specific average period of studies, we calculate weighted averages of the the average period of studies at universities and universities of applied sciences, respectively, using data from the Hochschulrektorenkonferenz (2012, p. 14ff). Accordingly, we assume that students enrolled at universities of applied sciences are eligible to four and students enrolled at universities are eligible to five years of BAföG funding. These cut-offs are rather restrictive to prevent an artificial increase in our NTU.

We abstain from further differentiating the maximum period of studies by desired degree for two reasons: First, we lack information on desired degrees and can only observe achieved degrees. Second, we have to rely on annual data for the students' enrollment status, so that we could not model the slight differences between the maximum period of studies in different degrees anyway.

Moreover, funding is granted on two further specific conditions. First, students have to proof sufficient progress in their studies. This proof of progress is due after completion of the fourth semester or, when their higher education institution requires taking an intermediate examination before the third semester, after completion of the second semester. Second, to remain eligible, students must not change their field of studies after a certain number

[27] We have to drop cases with missing information on the year of enrollment in higher education from our data.

of semesters. Lacking both information on grades and institution-specific information about intermediate examinations, we cannot incorporate the progress requirement. Because results from the representative student survey of Middendorff et al. (2013, p. 312) show that insufficient progress is neither important in students' decisions to apply nor a relevant factor to explain why students are not awarded the benefits, we consider this shortcoming as negligible. Lacking information on the students' field of studies, we can neither incorporate harmful changes in subject of studies. Changes in subject of studies are, however, also no frequent reason for why students forgo BAföG funding (Middendorff et al., 2013, p. 312).

As detailed previously, the family can protect additional monetary amounts for every sibling formally eligible to receive BAföG. To assess how many siblings are formally eligible to receive BAföG, we merge information on siblings from all survey waves of the SOEP. As the juridical distinction between eligible and ineligible educational programs is very complex and is often subject to individual case-by-case decisions, we cannot take into account all details of § 2 sect. 1 BAföG with the data at hand. We proxy siblings' eligibility, using information on their degree(s) previously attained, the type of their current educational program, and whether they are enrolled as full- or part-time students.

To set up the income tax model, we restrict our sample to eligible students for whom we have enough information to calculate the students' BAföG amounts. We face missings from three main sources: First, data on the students' wealth were only collected in 2002, 2007, and 2012. Second, data on parents' old-age provisions were only collected in 2004, 2006, 2007, 2010, and 2012. Third, data on church taxes payed were only collected in 2003, 2007, and 2011. As students' assets rarely exceed the allowable thresholds, missing information on assets are of minor importance for the quality of our calculation. Missing information on parents' old-age provisions and church taxes payed are more important because they directly affect the respective incomes relevant for the BAföG calculation. We follow Bruckmeier and Wiemers (2012) and linearly interpolate the missing values from all three sources for gap years. We have to discard cases where we do not even have enough information to interpolate.

To compute the incomes relevant for BAföG, we have to calculate the individual sum of all positive earnings as explained in section 3.9.1. We compute the sum of all positive incomes for each individual in the household, where possible. Income components such as profits or losses from investment of capital and rental income are, however, only available on the household level. We assume that these income components reduce or increase the income of the household head. As the income of married spouses enters the means test as an aggregate amount anyway and as only few cases report profits and losses at all, this assumption is innocuous for 94% of our sample.

The BAföG calculation uses parents' second last year's incomes as a default. If subsequent incomes are lower, e.g., because of unemployment, students can request using more recent incomes instead (see § 24 sect. 3 BAföG). We account for the possibility to update incomes by assuming that rational students request using parents' recent incomes if these are lower. Therefore, to compute the BAföG amounts between the years 2002 and 2013, we have to compute income taxes for the years 2000 through 2013.

Furthermore, we have to take into account income-related tax deductions from the parents' and students' incomes. Usually, these are considered up to a lump sum of EUR 1,000, unless higher expenses are proven for, e.g., commuting, moving, or working from home. We have enough information to calculate the most important part of the income-related tax deductions, namely commuting expenses. To calculate commuting expenses, we exploit available information on the commuting distance (single journey), the days worked (based on the annual working hours and taking into account information on full or part-time employment), and the deductible amount per kilometer in the respective year. We deduct the lump sum of EUR 1,000, unless the commuting expenses exceed EUR 1,000. In the latter case, we deduct the full commuting expenses.

Apart from that, we calculate further allowances for social security payments, but also income taxes, church taxes, and solidarity surcharges according to the respective German laws in the respective year (§ sect. 2 BAföG and German tax law (EstG)). The remaining BAföG calculation proceeds as detailed in the previous section.

All in all, the assumptions we have to impose on our microsimulation tend to underestimate parents' and, to a less extent, students' possibilites to protect income from the means test. Therefore, we tend to overestimate parents' financial resources available to support their offspring. In other words, our specification is rather restrictive. Restrictive assumptions are generally associated with a higher beta error and a lower NTU (Frick and Groh-Samberg, 2007). We discuss in the next section how relaxing the rather restrictive assumptions affects both measures.

3.9.3 Reduction of beta error

The students' level of needs is straightforward to calculate once the students' place of living, age, and family situation are known. Assuming that the microsimulation model correctly calculates the students' needs, there are two potential explanations for a high beta error rate: First, students are incorrectly classified as ineligible. As previously mentioned, our model tends to overestimate true incomes relevant for BAföG because we cannot incorporate all special allowances with the SOEP data. Accordingly, we tend to underestimate the number of eligible students which increases the beta error. Second, students are correctly classified as ineligible but their survey information on BAföG receipt is misleading.

The first case is plausible when the parents' or the students' incomes (as well as the students' own assets) exceed the respective thresholds only by a slight percentage. We study this possibility in models 1-4 of table A3.12. The first row contains the model used for our analyses as a benchmark. In model 1, for example, we consider students whose parents' relevant income or their own relevant income and assets exceed their needs by 5%. Assuming that students are classified as ineligible only because we overestimated their true incomes relevant for BAföG by up to 5%, we reduce our simulated incomes by the respective percentage and reclassify students from ineligible to eligible. Doing so makes our model less restrictive and decreases the beta error rate by 5.8% to 14.7%. The non-take-up rate is, however, very robust to this correction and decreases by 0.5% only. We report the sensitivity of beta error and NTU up to a correction of 20%. Although correcting incomes by 20% is a substantial reduction in BAföG-relevant incomes and makes our model far less restrictive, the non-take-up rate is only slightly affected. When compared to our benchmark model, the beta error rate decreases, however, by 20%.

In cases where the income relevant for BAföG exceeds the students' needs by far, but the student reports to have been funded, it seems more plausible that the information on benefit receipt is misleading. For example, in those cases where we cannot separate BAföG from merit-based scholarships, students can correctly report both positive student financial aid amounts and parents' incomes far beyond the respective BAföG thresholds. In models 5-8 of table A3.12, we investigate the sensitivity to reclassifying students from eligible to ineligible when the BAföG-relevant incomes are 10 times, 7.5 times, 5 times, and 2.5 times higher than the students' needs. As we observe less cases where the BAföG-relevant incomes exceed the students' needs by more than factor 10 than we observe cases where incomes exceed needs by only factor 2.5, model 5 has the least, model 8 the most impact on our beta error rate. The NTU is unaffected because we reclassify ineligible claimers to ineligible non-claimers, and both cases do not enter the NTU. Models 5-8 show that the beta error rate can be decreased to 9.9%, assuming that all cases with incomes exceeding needs by more than 2.5 times are in fact ineligible for BAföG.

Finally, models 9. - 12. combine both corrections. Reclassifying students as eligible if their (family's) BAföG-relevant incomes exceed their needs by up to 20% and as ineligible if their (family's) BAföG-relevant incomes exceed their needs by more than factor 2.5 leads to a decrease in the beta error rate by more than 50%. Nevertheless, the NTU is very robust even to these extensive corrections.

All in all, this analysis shows that the NTU is almost unaffected, although we allow for extreme and less realistic corrections of the incomes relevant for BAföG. As sensible corrections do also not affect our multivariate results, we decided to present results from the uncorrected model only (model 0).

Table A3.12: Sensitivity of NTU and beta error

Model	Beta error rate (%)	Non-take-up rate (%)
0. Reference	15.6	39.5
Correction if relevant income exceeds needs by up to		
1. 5 %	14.7	39.3
2. 10 %	14.1	39.1
3. 15 %	13.2	38.8
4. 20 %	12.5	38.7
Correction if relevant income exceeds needs more than		
5. 10 times	12.5	39.5
6. 7.5 times	11.4	39.5
7. 5 times	11.2	39.5
8. 2.5 times	9.9	39.5
Mixed		
9. model no. 1 and no. 5	11.6	39.3
10. model no. 2 and no. 6	9.9	39.1
11. model no. 3 and no. 7	8.7	38.8
12. model no. 4 and no. 8	6.7	38.7

Notes: The reference model is the specification used in our main analyses.

3.9.4 Additional tables

Table A3.13: Robustness of East German background effect

	(1) Probit		(2) Heckprobit	
	Coeff	AME	Coeff	AME
Simulated BAfoeG amount$^\natural$	−0.138**	−0.044***	−0.134**	−0.043***
	(0.054)	(0.017)	(0.052)	(0.016)
Age (centered)	−0.007	−0.002	0.013	0.004
	(0.034)	(0.011)	(0.034)	(0.011)
Female	−0.110	−0.035	−0.093	−0.030
	(0.143)	(0.046)	(0.138)	(0.045)
Migration background	−0.115	−0.036	−0.188	−0.060
	(0.206)	(0.064)	(0.199)	(0.062)
Living situation controls				
Student living in urban area	0.605***	0.189***	0.561***	0.178***
	(0.162)	(0.049)	(0.160)	(0.049)
Student living at parents' home	0.839***	0.268***	0.853***	0.276***
	(0.194)	(0.058)	(0.191)	(0.057)
Student lives in East Germany	0.252	0.081	0.275	0.088
	(0.232)	(0.073)	(0.232)	(0.073)
Parent and sibling controls				
Log parental current gross labor income$^\natural$	−0.030	−0.010	−0.013	−0.004
	(0.055)	(0.018)	(0.054)	(0.017)
Parent(s) have college degree	−0.127	−0.041	−0.108	−0.035
	(0.151)	(0.048)	(0.146)	(0.047)
Parents received social transfers	−0.043	−0.061	−0.058	−0.065
	(0.272)	(0.066)	(0.263)	(0.065)
East German background	−0.338	−0.139**	−0.409*	−0.162**
	(0.212)	(0.065)	(0.210)	(0.066)
Older sibling claimed BAfoeG	−0.702***	−0.210***	−0.735***	−0.224***
	(0.194)	(0.053)	(0.192)	(0.054)
East Germany × Social transfer last year	−0.572		−0.546	
	(0.420)		(0.411)	
Time-inconsistent preferences				
Willingness to take risks (std)	−0.102	−0.032	−0.100	−0.032
	(0.068)	(0.021)	(0.066)	(0.021)
Very impulsive	−0.101	0.028	−0.080	0.039
	(0.197)	(0.052)	(0.193)	(0.052)
Very impatient	0.002	0.067	−0.053	0.055
	(0.236)	(0.058)	(0.232)	(0.057)
Interaction effects				
Very impulsive × Very impatient	0.655*		0.707**	
	(0.349)		(0.341)	
Exclusion restriction (1st stage)				
Vocational training completed			−0.681***	
			(0.231)	
Observations	986		1041	
Baseline predicted probability	0.417		0.444	
corr(u1,u2)=ρ			−0.750	
Wald test ($\rho = 0$)			0.047	
Joint sig. of East German (p-value)	0.028		0.019	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: SOEP v30, 2002–2013, weighted. $^\natural$ = Deflated to base year 2007 and in hundreds of

Chapter 4

The Role of Information in the Application for Merit-Based Scholarships: Evidence from a Randomized Field Experiment

Stefanie P. Herber

4.1 Introduction

Student financial aid aims both at providing equal educational opportunities and at promoting the most talented students in higher education. While federal need-based aid emphasizes the goal to equate chances, federal merit-based aid focuses on promoting talents. Both forms of financial aid share the common feature that they are only effective if eligible students are aware of their existence and both willing and able to complete the complex paperwork involved when filing the application.

Regarding need-based financial aid, previous literature has built a case for information asymmetries and different levels of (parental) assistance between students of different socio-economic backgrounds (Scott-Clayton, 2013). A lack of information and assistance helps to explain why many eligible students of low socio-economic backgrounds do not file the complex application for need-based student aid (Dynarski and Scott-Clayton, 2006; King, 2006). Therefore, providing information and assistance can help diminishing this problem (e.g., Bettinger et al., 2012).

This chapter investigates whether information asymmetries are also relevant with respect to merit-based aid in Germany, where scholarship holders of non-academic backgrounds are considerably underrepresented (Middendorff et al., 2009).

In contrast to the US, where state scholarship programs are simpler, more transparent, and easier to apply to than need-based aid (Dynarski, 2004), or the UK, where the Higher Education Bursary and Scholarship Scheme assesses automatically whether the student qualifies for scholarships (Callender, 2009), the German system is much less transparent. Scholarships in Germany are tax-funded, but awarded by privately-owned providers in a highly competitive selection process to high-achieving applicants. The government only sets formal eligibility requirements, but leaves it to the 13 providers to define their own eligibility criteria. Consequently, the criteria vary extensively and are often not clear-cut. For example, most providers define no stringent grade point average needed to apply. For other selection criteria, such as certain personality traits, specifying cut-offs is impossible. This lack in transparency leaves room for information asymmetries, risks inefficient talent loss of high-achieving low socio-economic status students, and reinforces social inequalities. The latter is further accelerated by numerous non-monetary benefits from scholarships such as courses, personal support, and access to a social network of many high-profile alumni, which boost scholarship holders' careers after graduation.

In a randomized field experiment with over 5,000 German students, I study whether information asymmetries deter qualified students from applying for merit-based scholarships and whether mitigating these information asymmetries increases students' application rates. I randomly allocated participants to either the control or one of two treatment groups. In the first treatment group, participants received general, publicly available information on scholarships only. In the second treatment group, participants additionally received tailored information on the application process and probabilities of success, provided by a real, current scholarship holder. To ease identification, the scholarship holder resembled the participant in several characteristics, acting as a role model.

I consider two manifestations of information asymmetries that find expression in the design of the treatments.

First, prospective applicants must know the scholarship providers and their respective application requirements. As currently only 1% of all German higher education students are funded by these merit-based scholarships, compiling the distinctive details of the respective application procedures is challenging. This is especially true for students whose parents (and social surroundings) have not studied and cannot contribute their own experiences with student financial aid applications. The general information treatment addresses mainly this information asymmetry.

Second, potential applicants have to rate their own performance against that of their competitors in the selection process. Although all students face uncertainty about the sufficiency of their own qualification, non-academic students are disadvantaged in various ways. On the one hand, they can rarely benchmark their own performance against acquaintances who were successfully awarded a scholarship. On the other hand, the “cultural centeredness” (Steele et al., 2002, p. 420) of the German scholarship body reinforces the scholarship providers’ rather elitist appeal.[1] Consequently, high-achieving students of low socio-economic status might be afraid that stereotypes about “the educationally deprived” affect their chances to succeed (Steele et al., 2002, p. 422). The feeling not to fit into the scholarship system might put the applicants’ performance in the selection process under a stereotype thread or put students off applying entirely. A lack in role models of similar background to convey the credible assurance that students, who did not grow up in a family of academics, can be equally successful is then both cause and effect of a social selective scholarship system. The role model treatment considered here aims, first, at breaking this cycle to increase non-academic students’ sense of belonging to scholarship providers. Second, the role model treatment intends to provide insider information similar to that shared by parents or peers experienced with the German scholarship system.

My results provide evidence of information asymmetries and differences in previous applications between academic and non-academic students. Both treatments increased non-academic students’ poor baseline-knowledge about scholarships. Moreover, non-academic students in the role model treatment group doubled their applications for merit aid. Restricting the sample to the most eligible students increases role model treatment effects substantially. The general information treatment did, however, not affect application rates for merit-based scholarships—potentially because it triggered the students’ own information search for other, less selective, aid programs, and increased applications there. Furthermore, the treatments were ineffective for high-achieving female students who judge their own overall academic performance significantly less favorably than equally qualified men do.

This chapter adds to the existing literature in several ways. Up to now, little is known about whether information asymmetries between students of different socio-economic backgrounds matter also for high performing

[1] Translated literally, German scholarship foundations are promoting endowment, rather than providing aid on grounds of performance. Another example is that the Bavarian scholarship programs are regulated in the “Bavarian Elite Aid Act”.

students in higher education. To the best of my knowledge, this chapter is the first field experiment analyzing the effect of information provision with respect to merit-based scholarships and contributes to the sparse literature on information interventions in competitive settings. Furthermore, previous studies report mixed results as to whether the provision of information can indeed trigger behavioral changes and how interventions should be designed to do so. I shed further light on the design of interventions by testing whether participants lack information per se or tailored information provided by a similar role model. Finally, drawing on data on students' decision to apply, I can disentangle students' self-selection into the pool of potential scholarship holders from many other factors influencing whether they are indeed awarded the scholarship. I focus on students' applications because the scholarship providers' choice is limited to the pool of applicants. Therefore, from a policy perspective, equal opportunities at the stage of applications are the basis to secure an efficient and equitable allocation of funds.

The rest of this chapter proceeds as follows. After a review of the relevant literature in the next section, section 4.3 provides a short overview of the institutional background of merit-based student aid in Germany. Section 4.4 details the experimental set-up. Section 4.5 describes the data and gives brief descriptive analyses on heterogeneous information asymmetries and application experiences at baseline. Section 4.6 reports results of the experiment, and section 4.7 concludes. Supplemental tables, further details, and robustness checks can be found in the appendix.

4.2 Previous literature

Although I am unaware of experimental studies about merit-based student financial aid, numerous papers employ experimental set-ups to assess the behavioral impacts of information provision on costs and returns of going to college, or the availability of need-based student financial aid. Taken as a whole, evidence on the effectiveness of information interventions is mixed and depends both on the institutional context and the design of the intervention.

More specifically, a first strand of the literature intends to close information asymmetries by providing general information not customized to the recipient (e.g., statistics or leaflets on the returns to education). When official statistics are unavailable, not reliable, or poorly understood, a general information treatment can effectively increase years of schooling (Jensen, 2010), grades, and perceived returns to education (Nguyen, 2008) in developing countries or rural areas. In contrast to that, providing general information in industrialized countries cannot effectively increase take-up of student financial aid (Booij et al., 2012; Bettinger et al., 2012), college enrollments (Carrell and Sacerdote, 2013), or channel enrollment to degrees with higher educational returns (Kerr et al., 2014).

In industrialized countries with a broad coverage of publicly available information, customized information or personal assistance has a higher potential to affect behavior. Bettinger et al. (2012) study students' completion of the highly complex free application for federal student aid (FAFSA), which is central to access funds from most student aid programs in the US. The authors explicitly tested the advantages of providing personalized information and counseling over providing general information on student financial aid. They treated low-income students in all experimental groups with a brochure containing general information on costs and benefits of studying and need-based financial aid. The authors additionally provided one treatment group with individual aid estimates and encouraged them to file the FAFSA. Over and above both receiving general and personalized aid information, the third group was also offered personal assistance in completing the FAFSA. Only students in the personally assisted group were significantly more likely to receive aid, enroll, and persist in college.

Contrary to that, recent studies show that customizing information *can* positively affect low socio-economic status students' choice of more promising institutions or degrees (Hastings and Weinstein, 2008; Hoxby and Turner, 2013; Hastings et al., 2015).

With respect to coaching and counseling, Bettinger and Baker (2011) found that college students who received assistance in organizing their day and planning their studies were significantly more likely to persist and graduate. Likewise, Castleman et al. (2014) demonstrated in a recent study that counseling recent high school graduates on financial aid matters, reminding them of important deadlines, and assisting them with the paperwork increased their retention at and completion of college significantly.

Many other studies provide evidence that coaching or counseling increases the quality of educational choices and later labor market outcomes (e.g., Carrell and Sacerdote, 2013; Saniter and Siedler, 2014; Borghans et al., 2015), and might be even more cost-effective than increasing student financial aid (Bettinger and Baker, 2011).

Other studies maximize the targeting of the information by sending role models (or: peer counselors) with similar characteristics to students. Role models enhance the credibility of the information provided, increase the sense of belonging, and can induce participants to emulate them.

Nguyen (2008), for example, treated poor fourth-graders in Madagascar with three different interventions: A random group of students and their parents saw statistics on average educational returns to school only. The second treatment group met a person who shared his or her story of success with the children. Within this group, students were additionally randomly sampled to listening to the success story of a role model of high socio-economic background, or to a role model of shared, i.e., low socio-economic background. The third treatment group received both treatments. The author shows that both statistics and meeting a role model with shared characteristics can have large effects on perceived educational returns, attendance, and achievement of students of low socio-economic background. Combining both treatments increases, however, also awareness of the heterogeneity in educational returns and, therefore, reduces the positive effects of the statistics.

Dinkelman and Martínez A. (2014) take the same line with their intervention on low-income eighth-graders in Chile. They let students watch a 15-minute film where role models of similar socio-economic status describe financial aid possibilities. In consequence, the treatment increased students' high school enrollments and reduced school absenteeism.

Moreover, role models can be effective in stereotyped contexts such as math tests where women's ability (e.g., Marx and Roman, 2002) or at universities where non-academic students' performance is negatively stereotyped (e.g., Stephens et al., 2014). In these contexts, role models need not even share the stereotyped social identity (Steele et al., 2002, p. 428), though shared characteristics can increase the role models' effectiveness (Behncke et al., 2010; Marx and Ko, 2012).

In contrast to all that, the evidence on information asymmetries in competitive contexts such as applications for merit aid or at highly selective institutions is sparse. Yet, one study investigates talented low-income students' application behavior at selective US-colleges. Hoxby and Turner (2013) provided high-achieving students from low-income families with partly individualized, written information on the application process and personal expected net college costs at selective institutions. Furthermore, the intervention reimbursed treated students' application costs at up to eight colleges and also offered information for students' parents. Hoxby and Turner (2013) find economically and statistically significantly higher application and admittance rates to highly selective colleges. The mix of financial and informational incentives makes it, however, impossible to evaluate whether a gap in information or rather credit constraints were the decisive hurdle in students' access to selective higher education. Furthermore, Wiswall and Zafar (2013, 2015) show that even high-ability students at an elite university are not perfectly informed about returns from specific majors and that providing this information affects their choices. Unfortunately, the authors do not comment on heterogeneous effects by socio-economic background.

In sum, especially students of non-academic backgrounds should be more likely to show positive treatment effects if information is tailored and they can easily identify with a role model of similar socio-economic background. On the contrary, general information seems to be rather ineffective in impacting behavior. Up to now, we do, however, not know whether information asymmetries matter also for students who are of low socio-economic background, but score high in the achievement-distribution.

4.3 Institutional background

4.3.1 The German student aid system

As German colleges do not charge tuition, studying in Germany is relatively cheap in international comparison. Financial student assistance is, likewise, less pronounced when compared to countries charging high fees such as the US or the UK. Nevertheless, this means at the same time that German high schools usually lack a study adviser for student financial aid matters. Consequently, gaps in knowledge about how to finance studying persist.

Need-based income-contingent aid as of the Federal Training Assistance Act, short "BAföG", is the most common form of financial support in Germany, claimed by 17% of all enrolled German students in 2012 (German Bundestag, 2014). The state usually grants half of the BAföG amount as a subsidy, the other half as an interest-free loan. The loan-component must be repaid within 20 years after a grace period of five years. On average, funded students draw on a monthly funding amount of EUR 448 (Federal Statistical Office, 2015a, p. 32), which is equal to about 60% of the minimum subsistence level of a single person (German Bundestag, 2015, p. 8).[2]

The departments of the student services are responsible for counseling, processing of the students' applications, and calculating the respective funding amounts. These departments are closely associated with the respective higher education institutions, making BAföG a well-known funding source that students come across latest when they look for a room in one of the departments' student dormitories or charge their service cards for the canteens also operated by the student services.

[2] Therefore, most students have to rely on several financial resources. Therein, financial support by parents and own income from working besides the studies or in the semester break are most important (Middendorff et al., 2013, p. 593).

In contrast to that, the scholarship culture is still rather underdeveloped with currently not even 2% of all higher education students funded by some form of merit-based aid.[3] 13 privately-owned foundations for the promotion of young talent, called “Begabtenförderungswerke” (BFW), provide the most common form of merit-based aid in higher education. The foundations are privately owned and most of them pursue other goals over and above providing money for talented students, for example, political education, teaching of values in Germany and abroad, or development assistance. Therefore, and because the BFW are mainly funded by the German state, the merit-based aid system as a whole is obligated to reflect the plurality of society. Accordingly, each foundation is associated with a different facet of society: Several political foundations provide scholarships. Each of these foundations is affiliated with one of the parties in the German Federal Parliament. Moreover, there are religiously associated foundations and those affiliated with companies or trade-unions. Lastly, the ideologically neutral German National Scholarship Foundation is the oldest and largest BFW, promoting more than 40% of all funded scholars (German National Scholarship Foundation, 2014, p. 210).

The Federal Ministry of Education and Research is continuously extending funding amounts to increase the amount of scholarship holders. In 2014, EUR 232.6 million were provided to support 26,900 students enrolled in bachelor’s or master’s programs and 4,100 PhD students, summing up to about 1% of the overall student body (Federal Statistical Office, 2015b; Federal Ministry of Education and Research, 2015b). After the report of Middendorff et al. (2009) on the social selectivity in the German scholarship system spurred notable political and media attention (e.g. Kerbusk, 2009), special funds of EUR 8.2 million were placed at the BFWs’ disposal to increase the share of scholarship holders from “underrepresented groups”.

Unlike BAföG, both the BFW and the merit-based scholarships they award are completely separate of any (higher) education institution. This has two important implications. First, neither the amount nor the receipt of the scholarship is tied to visiting a certain university or being enrolled in a certain program. Second, the German merit-based aid scheme requires a high degree of the students’ own responsibility to get informed and to apply of their own accord to each BFW separately in order to participate in the respective selection processes.

Students usually apply for funding when they are enrolled in their first or second semesters of higher education, though some BFW allow applications for funding at the undergraduate level even before students officially enroll

[3] Own calculation based on Federal Statistical Office (2014b); Federal Ministry of Education and Research (2014a) and Federal Ministry of Education and Research (2015a) for 2013.

at university (tables A4.15 to A4.17 in the appendix give an overview). If the BFW also offers scholarships for Master's studies, students are usually required to apply before they start the program. From the respective pool of applicants, each BFW selects then its own future scholars (see the next section). When asked about acceptance rates, the BFW argue not to stick to a fixed rate but to admit all promising applicants.

Different to the US where students can claim both need- and merit-based aid simultaneously, German students have to decide between claiming need-based and merit-based aid. The latter is, however, clearly more favorable: Not only carry scholarships the advantage that they need not be repaid, they pay also higher aid amounts. Accordingly, the basic monthly scholarship awards are geared to the income-contingent BAföG amounts but supplemented by a monthly lump-sum amount of EUR 300. The resultant maximum award of EUR 970 is enough to concentrate fully on studying.

Beyond its financial advantages, a BFW scholarship signals high motivation and achievement of those who succeeded in the highly competitive selection process. A BFW scholarship is, therefore, considered a distinction worth being included in the curriculum vitae. Because the BFW aim at promoting and developing highly skilled young academics who are willing to take over social responsibility, funded scholars profit from many opportunities: The BFW provide conceptual support, such as interdisciplinary seminars, study trips, summer academies, personal support, and mentoring. With respect to their later career, funded scholars profit, moreover, from a rich alumni network which meets regularly and includes many high-profile politicians, researchers, and managers. Given that students of non-academic homes can draw on less financial resources and lack both counseling by college-experienced parents and a highly qualified network, they should benefit most from merit-based scholarships.

Apart from the most prominent form of merit-based aid provided by the BFW, a plethora of small private or institutional providers award scholarships to a small number of students. For example, some universities, companies, and cities provide merit-based scholarships to students born in the region or enrolled in a certain subject of studies. In comparison to the BFW, these scholarships are generally less focused on academic merit and impose more specific and more transparent criteria. As these scholarship providers are small and often only operating in a specific area, they are largely unknown and face far less competition. These scholarships are, therefore, potentially easier to win than the BFW scholarships (Pabst, 05.04.2015). Nevertheless, they do also usually pay less lucrative amounts than the BFW scholarships.

4.3.2 The application process for merit-based aid

The federal law only regulates that students are eligible to receive funding of the BFW "if their talent and personality promise outstanding performance during their studies and in working life" (Federal Ministry of Education and Research, 2014b, p. 3, own translation). They must furthermore meet some formal requirements, e.g., having a permanent residence permit and being enrolled full-time at a state-approved higher education institution. The further refinement of the aptitude criteria and the selection process is left to the discretion of each BFW.

Most BFW establish the following criteria to assess applicants' aptitude: First, applicants have to demonstrate "high performance" in high school or college. Second, applicants have to play an active part in society, politics, or culture, i.e., must be socially engaged, preferably compatible with the mission of the respective institution. Third, qualifying students must show responsibility, motivation, and dependability. Fourth, they should identify with the provider's alignment and goals, e.g., applicants at a Catholic BFW should identify with Catholic values. However, providers may put different emphases on the relative importance of these components and may also judge the "total package". Most BFW establish application thresholds with respect to acceptable age and semester ranges. Some BFW apply additional criteria, such as explicitly considering the applicant's socio-economic background. All in all, regulations and thresholds differ strongly between providers (tables A4.15 to A4.17 in the appendix give an overview).

Whether students meet the requirements to be funded during their studies is usually assessed in a very competitive procedure of several stages. For example, the German National Scholarship Foundation requires applicants to take an extensive test on their chances of academic success. After passing the aptitude test, they are invited to a selection seminar involving two interviews and a group discussion on short papers presented by the candidates. In 2013, 28.2% of the participants in the selection process were awarded a scholarship (German National Scholarship Foundation, 2014, p. 211).

The federal government explicitly supports the high heterogeneity in the application requirements and the selection processes to secure plurality in the scholarship body. Nevertheless, the resultant complexity increases the applicants' transaction costs to find an appropriate BFW. Because friends or parents with scholarships are much more common sources of information and motivation to apply than the high school or university,[4] non-academic

[4] A subsample of 376 participants in the experiment was funded by a BFW scholarship at wave 2. I exploited this coincidence by asking them additional questions after the general

students are more likely to lack important insights into the merit-based aid system. Accordingly, heterogeneous application requirements might equally well rather be detrimental to plurality.

Moreover, personality traits and volunteer work being core qualification requirements, it is impossible to define standardized eligibility cut-offs for sufficient qualification. Although academic merit should be easily quantified and compared, only a minority of BFW define a grade point average candidates must meet to successfully apply (grade point average (GPA) better than 2.0 on a five-point scale, 1.0 representing the best possible grade). In contrast to the transparent criteria underlying the provision of BAföG, students are highly dependent on forming expectations about their chances to succeed when applying for a scholarship.

4.4 The scholarship information experiment

The scholarship experiment was framed as a two-wave online survey on study finances with special focus on scholarships. I conducted the first survey between late October and early December 2013, and the second survey around half a year later (April/May 2014), i.e., in the first weeks of the winter and summer lecture periods, respectively.[5] To incentivize participation, students had the possibility to participate in a lottery which was tied to completing both waves.

4.4.1 Wave 1

For wave 1, I recruited participants via universities' official mailing lists where possible but also by means of printed posters and online study groups. The goal of the first survey was to gather information on the respondents' socio-economic and study background, to assess their knowledge of the German

part of the second survey, containing items on the sources that had informed them about the existence of BFW scholarships and the people who made them applying (multiple selections were possible). 36% were informed by friends and 22% by their parents, while 18% mentioned to have participated in an information program at their high schools or universities. Only 4% reported that an instructor at university or school had provided information on scholarships. More than half indicated that their parents had brought them to apply, 46% state that friends were the motivating factor. School teachers were named in 35% of cases and university lecturers in only 19%.

5 A third wave was conducted in May 2015, i.e., one year after the second wave, to give insights into whether students' scholarship applications were successful or not. Unfortunately, the response rate of students who applied after wave 1 was too small to conduct reliable analyses. Nevertheless, the data could be exploited to fill about 100 missings of time-invariant variables from wave 1, e.g., with respect to parental academic background.

scholarship system and to proxy their eligibility for a scholarship. Furthermore, I questioned participants on previous applications for scholarships.

After completing the questionnaire, respondents were randomly assigned to the control group or one of the two treatment groups. It is unsettled whether German students, especially freshmen, know of the rarely awarded scholarships at all and whether confronting them with potentially publicly available information on scholarships does already exert an effect. Therefore, I did not provide the control group with any general information. Along the same line of reasoning, I provided both treatment groups with a general information text to ensure their basic scholarship knowledge and the second treatment group's understanding of the role model interview.

More specifically, I randomly allocated participants to one of the following groups:

Control group: The control group was directly filtered to the last page where official university e-mail addresses were collected to invite participants for the second survey. As the universities' computing centers provide each student with a single university e-mail address once enrolled, I am able to restrict the sample to enrolled students and detect duplicates in my data.

General information treatment group: Participants were exposed to a text containing general information about merit-based scholarships, the amount of monthly funding, and formal application requirements. Text and graphics intended to offer objective information without explicitly encouraging students to apply. The wording was similar to an official website of the BFW Working Group (2013), especially when describing the respective application requirements. The text stressed, however, that students should gather more detailed information from the BFW directly.

Role model treatment group: The role model treatment group also received the general information text the information treatment group read, but was additionally provided with "custom-fit" insights through a written, personal testimony of a (real) student funded by one of the BFW.[6] I asked role models to answer a set of questions concerning personal benefits from scholarship and application requirements with a focus on the importance of academic achievement and social engagement. Role models should further detail the application and admission procedure, and estimate the chances to win a schol-

6 For the sake of credibility of and identifiability with the information and the scholarship holder, I decided to actually conduct interviews with 34 real scholars rather than confronting the participant with artificial vignettes. As I show later in the appendix, results are insensitive to potential slight variations between texts.

arship if belonging to a group currently underrepresented in the scholarship body. Although answers to these questions were tailored to the requirements of the specific BFW, all interviews shared a motivating tenor and stressed that an application, although strenuous, is worth the trouble. As the treatment focused non-academic students, role models also emphasized that students of non-academic backgrounds have equal chances to succeed and should not shy away from applying.

To avoid bad fit between role model and participant, e.g., a participant identifying with a left-wing party being matched with a scholar from a BFW associated with a conservative party, students were allocated to a role model based on their political and/or religious association. In order to achieve good matches, an algorithm (see appendix 4.8.1 for details) selected the interview which had the highest accuracy of fit with respect to field of studies and gender between the interviewed scholarship holder and the respondent. In other words, I established similarity on observed and controlled characteristics rather than additionally randomizing the degree of similarity.[7]

All interviews were headed with a warrant of apprehension (name, subject of studies, educational institution, semester, educational path to university) and showed the scholar on a casual photograph, so that participants could easily learn about the role model's characteristics.

4.4.2 Wave 2

Six months after the first survey, I invited approving students to access the second questionnaire via a personal link in their e-mail. The second survey aimed at updating information from the first survey, observing whether students' knowledge on scholarships changed, and refining judgment about their possible eligibility for a scholarship. Most importantly, I asked respondents whether they applied for a scholarship between both waves. As both personality traits and cognitive abilities are selection criteria for scholarships, the second survey included a short measurement of the Big Five Inventory BFI-S (Gerlitz and Schupp, 2005) and an untimed 12-item short-form of Raven's Advanced Progressive Matrices APM test (Raven et al., 1988), developed by Bors and Stokes (1998) and administered online.[8]

7 Slight differences in the quality of matching did not affect participants' application rates significantly (see tables A4.11 and A4.12 in the appendix).

8 In order to prevent attrition caused by an excessively long first wave, I decided to shift data collection of the BFI-S and the APM to wave 2. As the treatments are unlikely to affect measurement of personality and cognitive test scores and because both BFI-S and APM can be considered as relatively stable over time as indicated by acceptable test-retest stabilities (Hahn et al., 2012; Bors and Stokes, 1998), this should not affect the results.

4.5 Data

4.5.1 Descriptives

As students potentially eligible to receive a scholarship are the target group of this study, I restrict the sample to students enrolled in both waves. After removing 574 cases, including PhD students, recent graduates, and college drop-outs, 8,817 students who completed the first survey remained. Of these, 64.3% also finished the second interview.[9] Response rates for the second survey are also very similar between groups (controls: 65.0 %, info treatment: 64.2 %, role model treatment: 63.6 %) with differences between groups not being statistically significant (chi-squared test: $\chi^2 = 1.29$, $p = 0.53$). Moreover, using the wave 1 data set and regressing participation in wave 2 on the treatment dummies, the later baseline controls, and the interaction of both does not raise differential attrition concerns. Listwise deletion of participants with non-response on at least one of the items used as control variables (1.6% of the sample) results in a final analytic sample of 5,531 participants equally spread over groups.

In the final sample, participants study at about 180 different colleges (universities and universities of applied sciences), so that more than 40% of all German colleges are represented. I emphasize here, though, that the sample was not drawn on a representative basis as the population of students formally eligible to apply for a scholarship was unknown. The results and conclusions are, therefore, only internally valid.[10]

The following paragraphs describe the analytic sample and draw comparisons to the general student body, where possible. I focus on discussing means for the control group if not indicated otherwise.

Table 4.1 displays descriptive statistics within and between the three experimental groups. As shown in the last two columns, characteristics are balanced over groups, indicating that randomization was successful.[11] As

9 More than one third of wave 2 non-respondents (12.2% of those who finished wave 1) could not be contacted due to typos in the e-mail addresses collected. The high share of mistakes in e-mail addresses is probably due to the fact that most universities provide their students with randomly created, and hence hard to remember, addresses to prevent spam for and identification of the respective students.

10 Nevertheless, a self-selected sample, which is likely to represent the more committed students, is an appropriate potential target group for information campaigns of the BFW.

11 Members of the first treatment group were, however, marginally less likely to have applied for other scholarships ($p<0.1$). Applying procedures correcting for alpha inflation, e.g., Bonferroni-Holm, no statistically significant differences were found on an overall significance level of 1%. Figure A4.2 in the appendix shows that kernel density plots for the Big Five Inventory between groups are very similar.

is often the case in survey-based studies, female respondents are largely overrepresented, compared to official register data amounting to 48% female students (Federal Statistical Office, 2014c). More importantly, however, the representation by educational background is in range with recent representative student-level data reporting that 50% (*20. Sozialerhebung 2012* by Middendorff et al. (2013)) of the surveyed students are of non-academic background: Here, 52% of the participants are of non-academic background, meaning they descend from families where no parent achieved a college degree.

Compared to official register data for the sampling period (Federal Statistical Office, 2014a), students in my sample are more than 2 years younger (23 years, not reported) and have completed less semesters. I intended to sample students at an early stage of their studies as the BFW target students in their first semesters. Moreover, the students here are far more likely to be enrolled at a university (87% here vs. 58% in the register data). The overrepresentation of university students is a common phenomenon in survey data, even if drawn on a representative basis (and amounting to 74% in the *20. Sozialerhebung 2012*, for example).

Turning to key controls for the following analyses reveals that current holders of BFW scholarships (6%) are overrepresented as their share in the general student population amounts to only 1%. 16% of the students had already applied for a scholarship at a BFW, and 14% had applied elsewhere for a scholarship. Strikingly, the application rates by educational background differ only with respect to the BFW scholarships but not with respect to scholarships of other providers: While 14% of the non-academic students had applied at a BFW, the respective percentage of the academic students is more than one third higher (19%). At the same time, a similar proportion of academic and non-academic students had applied elsewhere (15% and 14%, $\chi^2 = 0.04$, $p = 0.84$). In contrast to the BFW, the non-BFW providers sometimes address students in financial hardship or of low socio-economic background directly, thereby likely to reduce information asymmetries by educational background. At the same time, however, many non-BFW providers also impose less challenging eligibility criteria, so that more non-academic students might qualify for other scholarships but not for the more selective BFW scholarships. In sum, this pattern already suggests that either a larger proportion of academic students is qualified to apply at a BFW or that non-academic students lack awareness of the profitable opportunities only BFW scholarships open up.

To proxy students' eligibility to receive a scholarship, the further analyses control for the fit of application requirements. As described above, dual degree students (12%), those studying in their second course of studies (4%), or part time (1%) are mostly ineligible to receive scholarships. Most providers require applicants to be at least younger than 35 years—which nearly all students in the sample are. Qualified applicants should officiate volunteer work (which half of the sample does) and show above-average academic performance. Because one third of all students in the sample were college freshmen in wave 1, they were not able to report grades of their studies yet.[12] Therefore, I used the study grades at baseline, where available, and substituted these by high school GPA if missing (2,010 cases).[13]

Representative data on students' average academic performance during their studies is unavailable in Germany. Taking information on high school GPA from the "mostly representative" (Ramm, 2014, p.10) *Studierendensurvey 2013* by Simeaner et al. (2014) as a benchmark suggests that high school GPAs in the sample here are overall similar but slightly better—which is reasonable because students in my sample are younger and average high school GPAs ameliorate continuously over cohorts.[14] About 45% of the sample here falls into the "high performance" group which is, according to the BFW that impose explicit study GPA-cutoffs, defined as a GPA better than 2.0 on the German five-point grading scale. 46% of the sample scores between GPA 2.0 and 2.9 (medium performance), only 9% scores lower than that. Despite the potential slight overrepresentation in study entrance grades, average cognitive test scores of 7.21 (S.D. = 2.70) are very close to results of the original offline version (mean = 7.15, S.D. = 2.34) used by Bors and Stokes (1998, p. 393).

As discussed earlier, delimiting the subsample with a viable chance to apply is difficult. Defining eligible students as students with high academic performance, who are younger than 35 years, neither dual degree students nor studying in their second course of studies, and have officiated volunteer work within the past 12 months, a share of 21.4% of this sample can be considered

12 There are also subjects of studies, e.g., Law, where the first semesters are not graded at all and grades are, naturally, missing.

13 This strategy should be unproblematic as students have to demonstrate their academic ability when applying for scholarships and will also have to use their high school diploma if they did not receive any college grades yet. Furthermore, if I used achievements as reported in the second semester, I would be unable to rule out bias introduced by potential treatment-related changes in achievement.

14 Data from the German Kultusministerkonferenz shows that high school GPAs ameliorated by roughly 2% between 2009 and 2013 over all German states. Grade inflation was most pronounced in North Rhine-Westphalia and Thuringia, where GPAs ameliorated by more than 5% (see figure A4.3 in the appendix).

as potentially eligible. This fraction reduces to 19% when I subtract current scholars. All these shares are equally spread over groups.

If not indicated otherwise, all analyses control for socio-economic and study-related characteristics, fulfillment of application requirements, the respective baseline levels of the dependent variable (applied at a BFW or applied at other non-BFW providers), and baseline scholarship receipt. Cognitive test scores and personality traits are added as indicated.

Table 4.1: Pre-treatment descriptive statistics

	(1) Controls		(2) General info		(3) Role model		(1) - (2)		(1) - (3)	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Diff.	(P-value)	Diff.	(P-value)
Female	0.68	(0.47)	0.68	(0.47)	0.67	(0.47)	0.00	(0.96)	0.00	(0.84)
Non-academic background	0.52	(0.50)	0.50	(0.50)	0.53	(0.50)	0.03	(0.12)	−0.01	(0.71)
Semester	4.43	(3.50)	4.39	(3.34)	4.53	(3.48)	0.04	(0.73)	−0.09	(0.42)
Type of institution										
University	0.87	(0.33)	0.87	(0.34)	0.86	(0.35)	0.00	(0.72)	0.02	(0.16)
Applied sciences	0.11	(0.31)	0.11	(0.31)	0.12	(0.33)	−0.00	(0.91)	−0.01	(0.22)
Other educational institution	0.02	(0.14)	0.02	(0.15)	0.02	(0.15)	−0.00	(0.55)	−0.00	(0.55)
Former receipt										
BFW scholarship wave 1	0.06	(0.24)	0.06	(0.23)	0.06	(0.24)	0.01	(0.52)	0.00	(0.76)
Other scholarship wave 1	0.04	(0.20)	0.03	(0.18)	0.03	(0.17)	0.01	(0.27)	0.01	(0.08)
Former application										
Applied at BFW	0.16	(0.37)	0.17	(0.37)	0.16	(0.37)	−0.01	(0.48)	−0.00	(0.71)
if non-academic	0.14	(0.34)	0.13	(0.34)	0.14	(0.34)	0.00	(0.80)	−0.00	(0.99)
if academic	0.19	(0.39)	0.20	(0.40)	0.20	(0.40)	−0.02	(0.32)	−0.01	(0.59)
Applied for other scholarship	0.14	(0.35)	0.12	(0.33)	0.13	(0.34)	0.02	(0.04)	0.01	(0.30)
if non-academic	0.14	(0.12)	0.13	(0.33)	0.13	(0.34)	0.02	(0.29)	0.01	(0.56)
if academic	0.15	(0.35)	0.12	(0.32)	0.13	(0.34)	0.03	(0.08)	0.01	(0.37)

Table 4.1: Continued

	(1) Controls		(2) General info		(3) Role model		(1) - (2)		(1) - (3)	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Diff.	(P-value)	Diff.	(P-value)
Application requirements										
Dual studies	0.12	(0.32)	0.11	(0.32)	0.10	(0.30)	0.00	(0.77)	0.01	(0.21)
Second degree	0.04	(0.19)	0.04	(0.19)	0.03	(0.18)	−0.00	(0.70)	0.00	(0.67)
Other non-eligible studies	0.01	(0.08)	0.01	(0.08)	0.00	(0.07)	−0.00	(0.83)	0.00	(0.37)
Older than 34 years	0.01	(0.10)	0.01	(0.08)	0.01	(0.11)	0.00	(0.23)	−0.00	(0.64)
Volunteer work	0.50	(0.50)	0.52	(0.50)	0.52	(0.50)	−0.02	(0.31)	−0.02	(0.36)
High performance	0.45	(0.50)	0.47	(0.50)	0.44	(0.50)	−0.02	(0.18)	0.00	(0.78)
Medium performance	0.46	(0.50)	0.44	(0.50)	0.46	(0.50)	0.02	(0.20)	−0.00	(0.90)
Low performance	0.09	(0.29)	0.09	(0.29)	0.09	(0.29)	0.00	(0.91)	−0.00	(0.79)
Cognitive abilities										
Cognitive test score	−0.01	(1.02)	0.02	(0.99)	0.01	(0.98)	−0.03	(0.40)	−0.02	(0.50)
Observations	1850		1839		1842		3689		3692	

Notes: The last two columns report p-values of independent samples t-tests. “Other educational institutions” include, e.g., teacher training colleges. Dual studies combine academic and practical phases financed by companies. Students in their second degree enrolled for a second undergraduate education after finishing their first undergraduate degree. Students in dual studies or their second degree are ineligible for most BFW scholarships. “Other non-eligible studies” include, e.g., part-time students. “Former application” indicates all applications up to wave 1 or current scholarship receipt in wave 1. “Other scholarships” includes scholarships such as the “Deutschlandstipendium” or company-provided scholarships. Cognitive test scores are standardized.

4.5.2 Application determinants

There are, of course, several reasons why students of non-academic backgrounds are underrepresented in the scholarship body. For example, a lower share of qualified students of non-academic backgrounds must translate into an equally reduced share in the overall scholarship body. Nevertheless, recent evidence suggests that differences in college grades between academic and non-academic students are very small (Delaney et al., 2011; Aspelmeier et al., 2012).[15] Even if the probability to meet the requirements was, however, unrelated to socio-economic characteristics, the selection process could introduce selectivity. College-experienced parents might, for example, coach their children to perform better, or students of non-academic background might perform worse when in a situation of stereotype threat.

Providing information can only exert an effect on a more equitable social composition if equally talented students of non-academic backgrounds are already underrepresented at the stage of applications. To explore whether this is indeed the case, I specify a logit model where I regress applications for a BFW scholarship up to the first survey on a set of socio-economic, college, and eligibility controls (table 4.2).

As expected, the application requirements are highly relevant determinants of the application decision with academic performance, volunteer work, and meeting the age requirement being most important.[16] Keeping all these factors at their observed values, students at universities of applied sciences are predicted to be about four percentage points less likely to report a previous application than students enrolled at universities. As the share of students who work besides their studies is higher in the applied sciences group, this effect is likely to capture more time constraints and a smaller financial need to apply for a scholarship.[17]

Furthermore, the results in column 1 suggest that respondents' socio-economic background influences application behavior. All else equal, the predicted probability to report an application was 2.5 percentage points (=18%) lower for students of families without academic experience than for students from academic homes. High achieving university students with

15 To the best of my knowledge, evidence for Germany is not available so far.

16 Of course, students in their second course of studies and students who are too old to be eligible may have applied earlier. The dummy flagging respondents older than 34 years does therefore also capture a time trend of scholarships being less frequent and known at the time they would have been eligible to apply.

17 Students' or parents' financial resources might be simultaneously affected by scholarship receipt (high income reduces the scholarship amount; scholarship funding increases financial resources). Lacking appropriate data, I cannot address this issue, unfortunately.

	(1)	(2)	(3)	(4)
Female	−0.016	−0.012	−0.023**	−0.021*
	(0.010)	(0.010)	(0.011)	(0.011)
Semester	0.004***	0.004***	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Non-academic background	−0.025***	−0.023**	−0.026***	−0.024**
	(0.009)	(0.009)	(0.009)	(0.009)
Applied sciences	−0.040***	−0.038***	−0.046***	−0.044***
	(0.014)	(0.014)	(0.014)	(0.014)
Other educational institution	−0.030	−0.032	−0.024	−0.026
	(0.030)	(0.029)	(0.030)	(0.030)
Medium performance	−0.157***	−0.152***	−0.143***	−0.137***
	(0.009)	(0.009)	(0.009)	(0.009)
Low performance	−0.162***	−0.160***	−0.155***	−0.153***
	(0.008)	(0.008)	(0.008)	(0.009)
Older than 34 years	−0.121***	−0.121***	−0.120***	−0.121***
	(0.030)	(0.030)	(0.030)	(0.030)
Dual studies	−0.036**	−0.033**	−0.033**	−0.030**
	(0.015)	(0.015)	(0.015)	(0.015)
Second degree	−0.028	−0.026	−0.032	−0.030
	(0.023)	(0.023)	(0.023)	(0.023)
Other non-eligible studies	−0.016	−0.013	−0.022	−0.017
	(0.064)	(0.065)	(0.061)	(0.062)
Volunteer work	0.166***	0.166***	0.162***	0.160***
	(0.009)	(0.009)	(0.009)	(0.009)
Party identification	0.028***	0.029***	0.023**	0.024**
	(0.010)	(0.010)	(0.010)	(0.010)
Cognitive test score		0.022***		0.025***
		(0.005)		(0.005)
Openness			−0.002	−0.003
			(0.005)	(0.005)
Conscientiousness			0.039***	0.040***
			(0.005)	(0.005)
Extraversion			0.007	0.011**
			(0.005)	(0.005)
Agreeableness			−0.012***	−0.013***
			(0.005)	(0.005)
Neuroticism			−0.005	−0.003
			(0.005)	(0.005)
Observations	5531	5531	5531	5531
McFadden's Pseudo-R^2	0.158	0.162	0.171	0.176

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Notes: Each column reports average marginal effects from a separate logistic regression on the probability that the participant had applied for a BFW scholarship at baseline. I conduct a principal component analysis and orthogonal varimax rotation (total explained variance = 65.28%) on the Big Five Inventory before extracting the five factors by regression scoring.

an average number of semesters (4.4), meeting all application requirements and reporting a party identification, had a 5.2 percentage points (=12.4%) lower predicted probability to have applied if of non-academic background ($p < 0.01$).

Omitted variable bias can, however, explain differences in applications if personality or cognitive abilities drive both application behavior and are correlated with socio-economic background. Therefore, I include covariates for cognitive test scores (col. 2) and personality traits (col. 3). It is well established that conscientious students who are likely to be motivated and to behave achievement-oriented perform better in college (e.g., O'Connor and Paunonen, 2007). Accordingly, I find that conscientious participants are predicted to be about four percentage points more likely to have applied (col. 3), over and above controlling for cognitive test scores (col. 4). Participants with high levels of agreeableness, being less assertive in their behavior, are predicted to be less likely, while extroverted individuals more likely to have applied when cognitive test scores are added. Nevertheless, none of the controls can close the application gap between students of different socio-economic backgrounds (col. 4).

What is striking, once personality traits are added, is that women are less likely to have applied. In the full specification, their predicted probability is 4.3 percentage points (10.9%) lower ($p = 0.05$) when considering eligible university students with average values on personality, test scores, and number of semesters.

Although I am not claiming causality here, the results provide some evidence that not only students of non-academic backgrounds but also women are already underrepresented when applying for scholarships, keeping eligibility constant. The lower application probability of women confirms the significantly smaller share of female scholarship holders in the German National Scholarship Foundation detected by Kuhlmann et al. (2012).

4.5.3 Information asymmetries

Is the decision to abstain from applying related to a lack in knowledge about scholarships? When asked about reasons for not applying, participants attach most importance to insufficient knowledge on application requirements, followed by insufficient volunteer work, and grades (table A4.14 in the appendix). Table 4.3 shows that students who had never applied at a BFW up to wave 1 were indeed poorly informed about scholarships. More than half of the participants indicated to be very or rather uninformed about scholarships, while only 9% stated to be informed or very informed.

Table 4.3: Knowledge level of non-applicants at baseline (wave 1)

	Mean	(S.D.)
Subjective knowledge level		
(Very) informed	0.09	(0.29)
Partly informed	0.36	(0.48)
(Very) uninformed	0.55	(0.50)
Knowledge on characteristics		
Amount correctly estimated	0.10	(0.30)
No provider known	0.36	(0.48)
Correct answer with respect to:		
Eligible even if not in upper half of very good grades	0.46	(0.50)
Own application without proposal possible	0.80	(0.40)
Amount need not be repaid	0.71	(0.46)
No strict grade requirements for prolongation	0.22	(0.42)
Knowledge indicator		
Sum of correctly answered	2.92	(1.35)
Observations	4622	

Notes: Participants who indicated not to know the answer to the question and those who failed to provide the correct answer were coded as 0, participants who came up with the correct answer were coded as 1. The "Knowledge indicator" sums participants' correct answers from all six objective knowledge items in the table.

This pattern is also found in participants' abilities to answer questions about scholarships correctly. Only 10% of the non-applicants were able to provide an estimate of the scholarship amount within an interval of EUR 50 around the true value of EUR 800.[18] Apart from that, more than one third could not name a single scholarship provider.

Several yes-no items tried to further assess students' perceptions of scholarships. Nearly half of the students knew that an application is possible without top margin grades. Most participants were informed about the possibility to apply at the BFW directly and knew that a scholarship need not be repaid. Yet, about 80% thought that a strict grade point average existed, which, if not met, led to a loss of funding.

In a nutshell, participants were inadequately informed and especially lacked knowledge on the flexibility of requirements. Summing up correct answers, respondents answered, on average, slightly less than half of the six items correctly. Less than 1% of the respondents answered all items correctly (not reported).

To explore information asymmetries, I regress the number of correctly answered questions on a set of controls, including eligibility requirements. To prevent reverse causality, I restrict the sample to those who had not applied at a BFW up to the first wave. The sample includes, therefore, both respondents totally unaware of scholarships and those who might have considered applying but decided against it. I run an ordered logistic regression, and, to facilitate interpretation, evaluate the results at the probability to answer five of the six questions correctly.[19] Table 4.4 reports average semi-elasticities. The average semi-elasticity indicates the average percentage change in the probability to answer five out of six questions correctly when the respective covariate increases by one unit, keeping all other variables at their observed values.

Unsurprisingly, academic achievement is, again, associated strongest with a high predicted probability of above-average knowledge: The predicted probability to answer five questions correctly is about 66 (116) percent lower for participants with moderate (low) instead of high academic achievements, ceteris paribus. Dual study students usually ineligible to receive scholarships are predicted to be 21 percent less likely to provide five correct answers.

---

[18] Respondents were asked to name the scholarship amount equivalent to EUR 500 of BAföG. Respondents therefore needed to know that the scholarship amount equals BAföG, but that scholarship holders receive a lump-sum payment of EUR 300 on top.

[19] Estimates across all other cut-offs are shown in the appendix exemplary for the specification of column 1 in table 4.4 (see figures A4.4 and A4.5 in the appendix). Patterns for the other specifications are similar.

Older students tend to be better informed, possibly because they had more opportunities to meet scholarship holders during their studies in comparison to young students. The socially engaged who are more likely to meet funded scholars during volunteer work, are predicted to be about 14 percent more likely to answer five questions correctly.

The predicted probability of above-average knowledge for a non-academic at university with an average number of semesters, meeting all eligibility requirements is about 16% ($p < 0.01$) lower than that of a comparable student with college-educated parents. Calculating the same average semi-elasticity with respect to gender, women's predicted probability is about 26% ($p < 0.01$) percent lower than that of similar men. These results are only slightly affected when controlling for potential differences in cognitive abilities in column 3.

To explore in how far this effect is mitigated by informal knowledge within the social network, I add a dummy for acquaintances with a scholarship holder (col. 2). People who indicated to know a (former) scholarship holder had substantially higher predicted probabilities to be informed. As significantly less non-academic students were acquainted with a scholar than their counterparts from academic homes ($\chi^2 = 59.16$, $p = 0.00$), the difference in knowledge between academic and non-academic students drops by about one quarter but is not completely offset. Note that the inclusion of the informal knowledge dummy does not affect the gender gap. As I cannot reject the hypothesis that men and women differ in their probabilities to know a scholar ($\chi^2 = 0.86$, $p = 0.35$), the results suggest that information asymmetries might be a relevant obstacle for non-academic students but probably not for women. The same holds for including cognitive test scores in column 4.

In column 5, I isolate the effect for those who had actively looked for information but then decided against applying by adding a dummy on own information search.[20] The influence of grades, volunteer work, and dual studies is reduced, indicating that the most eligible did indeed inform themselves and were thus better prepared to answer the questions. Strikingly, gaps with respect to academic background and gender increase, emphasizing that non-academic and female students were also less likely to have looked for information.

[20] As only respondents who had not applied for scholarships and were not planning to do so at baseline received this question, the sample size reduces to 2,681 students.

Table 4.4: Ordered logit model for scholarship knowledge: average semi-elasticities

	(1)	(2)	(3)	(4)	(5)
Female	−0.289***	−0.290***	−0.260***	−0.263***	−0.324***
	(0.048)	(0.049)	(0.049)	(0.049)	(0.070)
Semester	0.031***	0.021***	0.029***	0.020***	0.017*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)
Non-academic background	−0.177***	−0.138***	−0.165***	−0.127***	−0.287***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.066)
Applied sciences	−0.051	−0.046	−0.032	−0.029	−0.188*
	(0.073)	(0.073)	(0.073)	(0.073)	(0.102)
Other educational institution	−0.038	−0.018	−0.064	−0.042	−0.185
	(0.166)	(0.167)	(0.168)	(0.168)	(0.243)
Medium performance	−0.657***	−0.643***	−0.629***	−0.617***	−0.346***
	(0.050)	(0.050)	(0.050)	(0.050)	(0.076)
Low performance	−1.155***	−1.089***	−1.116***	−1.054***	−0.666***
	(0.084)	(0.085)	(0.085)	(0.085)	(0.108)
Older than 34 years	0.471**	0.474**	0.470**	0.472**	0.290
	(0.185)	(0.185)	(0.185)	(0.185)	(0.298)
Dual studies	−0.209***	−0.202***	−0.199***	−0.192***	−0.126
	(0.072)	(0.072)	(0.073)	(0.072)	(0.101)
Other non-eligible studies	0.330	0.408	0.379	0.450	0.686
	(0.363)	(0.370)	(0.357)	(0.363)	(0.502)
Volunteer work	0.141***	0.096**	0.140***	0.096**	−0.007
	(0.045)	(0.046)	(0.045)	(0.046)	(0.066)
At least one acquaintance		0.449***		0.439***	0.301***
		(0.050)		(0.050)	(0.068)
Cognitive test score			0.125***	0.117***	0.134***
			(0.024)	(0.024)	(0.034)
Actively looked for information					1.086***
					(0.076)
Observations	4622	4622	4622	4622	2671
Baseline predicted probability	0.039	0.039	0.039	0.039	0.057
P-value overall Brant test	0.775	0.688	0.624	0.530	†

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: The table shows average semi-elasticities from an ordered logit model. Average semi-elasticities are calculated for the probability to answer five of the six items correctly. Figures A4.4 and A4.5 in the appendix show how average semi-elasticities vary over cut-offs. The sample is restricted to those who had not applied for a scholarship up to wave 1. †=Too few observations in some sub-groups to compute the Brant test.

4.6 The effects of information provision

4.6.1 Method

In the following, I analyze intent-to-treat (ITT) effects for five dependent variables, all of them measured at the time of the second survey: the number of correctly answered knowledge items, whether participants applied for a merit-based scholarship at a BFW or at other providers, whether they thought about applying, and, finally, whether they actively engaged in gathering more information about scholarships.

I estimate all ITT effects by specifying the following model:

$$y_i = \beta_0 + \beta_1 \cdot \mathit{INFORMATION} + \beta_2 \cdot \mathit{ROLE\ MODEL} + \boldsymbol{x}_i^{'} \cdot \boldsymbol{\beta}_3 + \epsilon_i, \quad (4.1)$$

where y_i is the respective outcome variable for student i at the time of the second survey. The treatment dummies *INFORMATION* and *ROLE MODEL* indicate whether students received only general information, or whether they received the role model treatment. β_1 and β_2 represent the intent-to-treat effects of the information and role model treatment, respectively. As previously mentioned, the role model treatment also included the information treatment. Therefore, β_2 represents the composite effect of both treatments with respect to no treatment. $\boldsymbol{x}_i$ is a vector of baseline controls. ϵ_i represents the error term, estimated using robust standard errors.

When the dependent variable is the number of correctly answered knowledge items, I omit the constant from equation 4.1 and estimate an ordered logit model. y_i becomes a latent variable of the students' scholarship knowledge, observed in one of the seven categories from zero items to six items answered correctly. For the ease of interpretation and similar to the previous analyses of students' information asymmetries (table 4.4), I report average semi-elasticities for the probability to answer five of the six knowledge items correctly.

For all other binary dependent variables, I run simple linear probability models with ordinary least squares (OLS); non-linear specifications yield, however, similar results.

Being interested in non-academic students' application probabilities mainly, I investigate heterogeneous ITT effects by adding interactions between the treatment dummies and the students' educational background in most of the analyses.

4.6.2 Results

Treatment effects on scholarship knowledge

Table 4.5 shows whether the treatments increased scholarship knowledge at the time of the second survey for the whole sample (col. 1–2) and by educational background (col. 3–6). Columns 1 and 2 indicate that both treatments increased the knowledge about scholarships significantly.[21] The average predicted probability to answer five of the six knowledge questions correctly increased by about 12 percent in both the information and the role model treatment group (col. 1). Adding personality traits and cognitive test scores slightly decreases the estimates. Decomposing the sample by educational background reveals that the effects are twice as large and only statistically significantly different from zero for non-academic students for whom the information was designed and who were worse informed at baseline.

Table 4.5: ITT effects on knowledge: Ordered logit model

	All		Non-academic		Academic	
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	0.120**	0.111**	0.146**	0.142**	0.089	0.073
	(0.047)	(0.047)	(0.067)	(0.067)	(0.066)	(0.067)
Role model treatment	0.119**	0.113**	0.159**	0.157**	0.072	0.060
	(0.048)	(0.048)	(0.067)	(0.068)	(0.068)	(0.068)
Big5 Controls		✓		✓		✓
Cognitive test scores		✓		✓		✓
Observations	5195	5195	2726	2726	2469	2469

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: ITT effects reported as average semi-elasticities from an ordered logit model (cut-off at 5 correctly answered items) with respect to the control group. Each estimation controls for all covariates of table 4.1, including the baseline level of the respective dependent variable.

Treatment effects on applications for scholarships

This section investigates whether the better knowledge about scholarships carried over to students' higher application rates at wave 2. Table 4.6 presents

[21] Results without covariates for this and all following specifications are very similar. I report covariates-adjusted results only as these are more efficient and take care of potential remaining differences between groups. Unadjusted results are available upon request.

the impact of both treatments on applications at a BFW (col. 1–3) and at other scholarship providers (col. 4–6). As a reference point, the bottom of the table contains the predicted probabilities to apply for a scholarship at wave 2 for the control group. I start with discussing the treatment effects for applications at a BFW.

Considering the whole sample, column 1 reveals that less than three percent of the students applied for a BFW scholarship between wave one and wave two. Moreover, neither of the two treatments had a statistically significant effect on applications.

Decomposing the sample again by educational background shows that the role model treatment increased non-academic students' application rates by highly statistically significant 2 percentage points (col. 2–3), though. In other words, in comparison to the respective control group, where only 1.9% of the students with non-academic background applied for a BFW-scholarship, the role model treatment more than doubled non-academics' application probabilities. In contrast to that, the respective treatment effect is not statistically significant for students of college-educated families ($p > 0.1$). Although non-academic students in the general information treatment group are equally likely to answer objective knowledge questions on scholarships, their probability to have applied for a BFW scholarship is not significantly affected (col. 2–3) and marginally significantly smaller than the ITT of the role-model treatment ($p < 0.1$). Therefore, only the role model treatment increased a sense of belonging, allowed students to look behind the scenes, and to accumulate insider information relevant to decide in favor for applying themselves. As a consequence, ITT effects found in the second treatment group can be considered as stemming from the interview text and not from the general information text which was also provided to the general information group.

Turning to students' application probabilities for other scholarship opportunities not provided by the BFW (col. 4–6 in table 4.6) shows that the effect pattern between both treatment groups reverses. Participants in the general information group were 1.5 points more likely to report applications for non-BFW scholarships, though the effect is only marginally statistically significant. At the same time, the coefficient for members of the other treatment group is negligibly small and insignificant, though not significantly smaller ($p > 0.1$, col. 4). The negative signs of the interactions with educational background point to a smaller effect for academic students, yet, the interactions are not significantly different from zero (col. 5-6). This is in line with the observation that baseline application rates for non-BFW scholarships were already very similar for academic and non-academic students (see table 4.1).

Table 4.6: ITT effects for full sample and heterogeneous effects: OLS

	Application at a BFW			Application elsewhere		
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	0.001	0.007	0.007	0.015*	0.019*	0.019*
	(0.005)	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
Role model treatment	0.005	0.020***	0.020***	0.002	0.002	0.003
	(0.006)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)
Academic background	−0.004	0.010	0.010	0.004	0.006	0.007
	(0.005)	(0.008)	(0.008)	(0.007)	(0.011)	(0.011)
Interaction effects						
Info × Academic		−0.012	−0.012		−0.007	−0.007
		(0.011)	(0.011)		(0.016)	(0.016)
Role model × Academic		−0.032***	−0.032***		0.000	−0.001
		(0.011)	(0.011)		(0.015)	(0.015)
Big5 Controls			✓			✓
Cognitive test scores			✓			✓
Observations	5195	5195	5195	5195	5195	5195
Pred. probability to apply between W1 and W2 (control group)						
All	0.026	0.026	0.026	0.055	0.055	0.055
Non-academic		0.019	0.019		0.050	0.050
Academic		0.034	0.034		0.060	0.060

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: ITT effects reported with respect to the control group. Each estimation controls for the covariates of table 4.1, application at baseline and the receipt of other scholarships at baseline. Additional covariates or interactions with both treatment groups are added as indicated. 334 participants were dropped because they were already funded by a scholarship (parallel funding not possible).

Treatment effects on short-run behavior

To better understand why the general information treatment increased participants' applications for non-BFW scholarships but did not affect applications for BFW scholarships, I investigate whether the treatments had similar effects on intentions to apply and own information search.[22] Table 4.7 contains regression results for the dependent variables having thought about applying for any scholarship (col. 1–2) and having actively looked for more information after the treatment (col. 3–4).

In the control group, about 42% of the students had thought about applying at the time of the second survey. Both treatments increased the share of those who had thought about an application by a small but significant amount. Although both treatments had an equally large impact on whether participants considered applying ($p > 0.1$, col. 1–2), only the general information treatment triggered own active information search between wave 1 and wave 2. The effect for participants in the role model treatment group who were provided with extensive information is significantly smaller ($p < 0.1$, col. 3) and not statistically significant overall. That the treatment effect in the information treatment group is not statistically significantly different from zero is not surprising as the general text provided only basic information about scholarships and urged participants to go online to look for more extensive information. In this vein, students in the information treatment group might have come across other, probably more suitable or less challenging, scholarship opportunities and applied there, while those treated more extensively restricted their attention to applying at a BFW. As outlined previously, the selection criteria of other, smaller scholarship programs are often more transparent and have less competitive and complex selection processes. Publicly available information is, therefore, more helpful when gathering information on non-BFW scholarships than when cutting one's way through the more complex information on scholarships provided by one of the BFW.

Treatment effects on applications of the most eligible students

Because the majority of the students in my sample are not eligible to receive scholarships, the treatment effects presented so far are only a lower bound of the true effects. To purge the sample of mostly ineligible students, table 4.8 restricts the sample to an approximation of the target population. I start

[22] Unfortunately, only those who had never applied for any scholarship at baseline and did not intend to do so before the treatment were asked whether they had actively looked for information or thought about applying.

Table 4.7: ITT effects on pre-application outcomes: OLS

	Thought about applying†		Active information search†	
	(1)	(2)	(3)	(4)
Information treatment	0.059***	0.057***	0.036**	0.034*
	(0.020)	(0.020)	(0.018)	(0.018)
Role model treatment	0.044**	0.044**	0.004	0.006
	(0.020)	(0.020)	(0.018)	(0.018)
Big5 Controls		✓		✓
Cognitive test scores		✓		✓
Observations	2670	2670	2670	2670
Pr(y=1) at W2 (control group)				
All	0.419	0.419	0.233	0.233

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: ITT effects reported with respect to the control group. Each estimation controls for all covariates of table 4.1, including the baseline level of the respective dependent variable. † These items were only given to respondents who had not applied for scholarships and were not planning to do so at baseline.

by dropping all students likely to be ineligible to apply at most BFW because they are too old or study in ineligible programs (col. 1–2). Then, I also drop students beyond the most favored range of semesters (col. 3). In column 4, I exclude moderately and low performing students and in column 5 also those who exert no volunteer work. Columns 6 and 7 keep, in addition to column 3, only the high performing students. I depict results without educational background interactions for the first sample reduction (col. 1) and the most restrictive sample (col. 6).

Columns 1 and 6 without interactions reveal that both treatments remain insignificant in the sample of most eligible students. As students in the general information group applied at higher rates elsewhere, irrespective of their factual eligibility to receive BFW scholarships,[23] ITT effects in the general information group do not increase between columns 1 and 6 and stay negligible in size and statistical significance. Although the ITT effects in the role model group increase up to 2.2 points in column 6, the effect is still not statistically significant.

Contrary to that, heterogeneous ITT effects for non-academic students are statistically significant in all specifications (col. 2–6, col. 7) and rise steadily up to 6.6 percentage points (col. 7). In the specification including only students with volunteer work and high academic performance (col. 5), application rates for non-academic students in the role model group are about twice as large as the respective control group benchmark. In the most restrictive specification for formally eligible, high-achieving students in the eligible semester range (col. 7), application rates in the role model treatment group exceed the respective control group benchmark by more than factor 5. As can be seen from the bottom of the table, the predicted probabilities for the control group to have applied for a BFW-scholarship differ significantly between non-academic (row b) and academic students in all specifications (row c): While academic students in the control group were increasingly more likely to have applied for a scholarship if they are highly eligible, even highly eligible non-academic students in the control group were up to 7 percentage points less likely to have applied. The role model treatment closed the gaps in application probabilities between students of different educational backgrounds almost completely.

[23] Students ineligible for BFW scholarships have *not* applied less often for other alternatives when I repeat the analysis from table 4.8 but restrict the sample up to the least eligible participants. ITT estimates are similar to the unrestricted sample and larger in the information treatment group.

Table 4.8: ITT effects of application for a BFW-scholarship, approximation of the relevant sample: OLS

	Formal criteria		Formal crit.+Semester	Performance	Volunteer	Performance+Semester	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Information treatment	−0.004	0.004	0.019	0.013	0.005	0.005	0.057*
	(0.006)	(0.007)	(0.016)	(0.016)	(0.024)	(0.024)	(0.031)
Role model treatment	0.002	0.019**	0.027*	0.034*	0.065**	0.022	0.066*
	(0.006)	(0.008)	(0.016)	(0.018)	(0.030)	(0.026)	(0.034)
Academic background	−0.005	0.013	0.026	0.038**	0.048*	0.011	0.071**
	(0.005)	(0.008)	(0.018)	(0.018)	(0.028)	(0.020)	(0.032)
Interaction effects							
Info × Academic		−0.017	−0.033	−0.044*	−0.056		−0.095**
		(0.012)	(0.025)	(0.024)	(0.036)		(0.046)
Role model × Academic		−0.037***	−0.032	−0.066**	−0.102**		−0.081
		(0.012)	(0.027)	(0.026)	(0.041)		(0.052)
Observations	4232	4232	1182	1781	896	550	550
Pred. probability to apply between W1 and W2 (control group)							
(a) All	0.026	0.026	0.030	0.052	0.065	0.051	0.051
(b) Non-academic		0.018	0.014	0.030	0.036		0.012
(c) Academic		0.036	0.047	0.071	0.090		0.086
P-value (c)=(b)		0.036	0.060	0.026	0.055		0.020

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: ITT effects reported with respect to the control group. Each estimation controls for socio-economic characteristics, application requirements, former applications, and current receipt of other scholarships than BFW scholarships. Starting from the full sample and excluding current scholarship holders of wave 1, the sample is further restricted as follows: “Formal” = drop dual degree, part-time, students in their second course of studies, students older than 34 years in wave 1, and current scholarship holders; “Formal+Semester” = additionally drop students in their Bachelor’s but higher than in second semester at baseline or in their Master’s and higher than first semester in wave 2; “Performance” = in addition to “Formal”, keep only high performing students; “Volunteer” = in addition to “Performance”, keep only socially engaged students; “Performance+Semester” = in addition to “Formal+Semester”, keep only high performing students.

Success of applications

How likely is it that the non-academic students who applied after the treatment are indeed awarded the scholarship? Unfortunately, I cannot assess this question directly because the response rate of applicants assessed in a third wave 1.5 years after the first was too low to conduct meaningful analyses.

Nevertheless, I can investigate whether the success probabilities of students who had applied before wave 1 differed by educational background. Table 4.9 reports OLS-estimates for the subsample of those who reported a BFW-application at baseline. More specifically, I regress an indicator of having applied (un-)successfully on several covariates. The estimates show that non-academic students were not statistically significantly less successful than comparable academic students.

These results are confirmed when I broaden the focus to applications up to wave 3. I do not find evidence that students of non-academic background who applied for the scholarship after wave 1 have a different likelihood to be awarded the scholarship ($\chi^2 = 0.43$, $p = 0.51$).[24]

Furthermore, I asked students who were already financed by a BFW-scholarship at wave 1 to assess, in how far the probability to be awarded a scholarship is affected if applicants are of low socio-economic background (but, other than that, similar to all other applicants). Respondents could choose between the categories “very negatively”, “rather negatively”, “no influence”, “rather positively”, and “very positively”. 76% assumed that the probability to be awarded the scholarship is rather or very positively affected if applicants are of low socio-economic background, ceteris paribus. Answers to this question differed not significantly by respondents’ own socio-economic backgrounds ($\chi^2 = 3.37$, $p = 0.50$). This assessment is also in line with the stated goal of the BFW to increase the share of underrepresented students.

All these results confirm the descriptive findings of Kuhlmann et al. (2012) who report no statistically significant differences in general acceptance rates by (non-)academic background for the German National Scholarship Foundation and even higher acceptance rates for first-semester students without parental college degree. Taken together and against the background that additional funds to increase the percentage of underrepresented groups in the merit-based aid system are available, a higher number of qualified

24 The likelihood to participate in the third wave differed not significantly between students of non-academic and academic background ($\chi^2 = 0.12$, $p = 0.72$), and both groups were equally likely to participate in wave 3 if successfully awarded a scholarship ($\chi^2 = 0.67$, $p = 0.41$).

applicants of non-academic background should most likely also translate into a higher number of scholarships awarded to them.

Table 4.9: Success probabilities for applications at baseline: OLS

	(1)		(2)		(3)	
Socioeconomic background						
Female	−0.034	(0.031)	−0.031	(0.031)	−0.009	(0.031)
Semester	0.074***	(0.013)	0.067***	(0.012)	0.066***	(0.012)
Semester^2	−0.003***	(0.001)	−0.003***	(0.001)	−0.003***	(0.001)
Non-academic background	−0.035	(0.030)	−0.029	(0.029)	−0.023	(0.029)
Type of institution						
Applied sciences	0.044	(0.059)	0.030	(0.058)	0.042	(0.056)
Other educational institution	0.110	(0.117)	0.084	(0.110)	0.100	(0.112)
Academic performance						
Medium performance	−0.065*	(0.036)	−0.073**	(0.035)	−0.074**	(0.036)
Low performance	−0.159	(0.104)	−0.220**	(0.099)	−0.231**	(0.095)
Volunteer work			0.268***	(0.027)	0.248***	(0.027)
Personality traits						
Openness					0.012	(0.014)
Neuroticism					−0.062***	(0.015)
Agreeableness					0.020	(0.014)
Extraversion					0.029*	(0.015)
Conscientiousness					0.012	(0.015)
Cognitive abilities						
Cognitive test score					0.014	(0.015)
Observations	897		897		897	
McFadden's Pseudo-R^2	0.048		0.095		0.107	
Predicted probability of success			0.299		0.299	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: The sample is restricted to students who mentioned to have applied at a BFW at baseline. The dependent variable equals 1 if the student received the scholarship as reported at baseline and equals 0 if the student did not receive the scholarship.

Differences between men and women

To follow up briefly on the gender effects I found with respect to information levels and former applications, I investigate in table 4.10 whether these effects carry over to gender-differences in treatment effects.[25] Surprisingly, the role model treatment did not affect women's application probabilities for merit-based scholarships significantly, whereas the effects are statistically significantly different from zero for men. This finding corroborates once more that information asymmetries are not key to explain why female students apply less often (see section 4.5.2) and are less likely to persist in the selection process of the German National Scholarship Foundation, although equally eligible (Kuhlmann et al., 2012).

Table 4.10: ITT effects of application for a BFW-scholarship, by gender: OLS

	(1)		(2)		(3)	
Information treatment	0.001	(0.005)	−0.000	(0.007)	0.000	(0.007)
Role model treatment	0.005	(0.006)	−0.003	(0.007)	−0.003	(0.007)
Male			−0.006	(0.008)	−0.004	(0.008)
Interaction effects						
Info × Male			0.003	(0.011)	0.003	(0.011)
Role model × Male			0.026**	(0.013)	0.026**	(0.013)
Observations	5195		5195		5195	
Pred. probability to apply between W1 and W2 (control group)						
(a) All	0.026		0.026		0.026	
(b) Female			0.027		0.027	
(c) Male			0.023		0.023	
P-value (c)=(b)			0.618		0.617	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

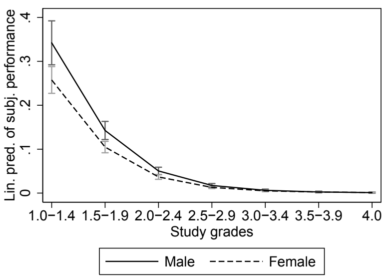
Notes: ITT effects reported with respect to the control group. Each estimation controls for the covariates of table 4.1, application at baseline and the receipt of non-BFW scholarships at baseline. Additional covariates or interactions with both treatment groups are added as indicated.

One possible explanation for the finding that equally qualified women abstain from applying might be the underestimation of their own abilities and the lower level of confidence about their own performance (e.g., Deaux and Farris, 1977; Chevalier et al., 2009). Another explanation is women's generally higher average performance in college (Vincent-Lancrin, 2009). If women compare their own achievement to that of their peer group, their

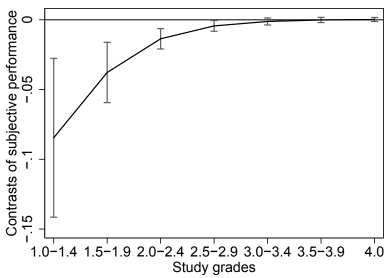
[25] The results for a sample restriction similar to table 4.8 are even more pronounced and yield zero effects for women but larger effects for men. Because sample sizes for men become relatively small when cutting down the sample, the results (available on request) have to be interpreted with caution, though, and are therefore not reported here.

self-assessment might be lower just because the average level of performance in a female-dominated peer group is higher.

To investigate these channels, I run an ordered logistic regression of participants' self-assessed academic performance with respect to their peers on several covariates: gender, study grades, field of studies, a three-way-interaction of gender, study grades, and field of studies plus their composite terms, cognitive test scores, high school GPA, type of institution, semester, and personality traits. I cluster the standard errors at the respective higher education institution to account for differences in grading and quality between universities. Keeping the factual study grades constant, I find that men are significantly more likely than women to evaluate themselves as "much better" than their peers if their study GPA is better than 3.0 (see figure 4.1). While the average predicted probability of men with top grades to evaluate their own performance as "much better" amounts to 32%, the respective average probability of women is about one third lower. Lacking objective measures for the peer groups' performance, I cannot assess, however, whether it is women's lower self-consciousness or their different reference point that drives their worse self-assessment.



(a) Linear predictions



(b) Contrasts of linear predictions

Figure 4.1: Linear predictions of students' subjective performance, evaluated against their peers

Notes: Results from an ordered logit regression. Dependent variable: own study GPA is much worse (1), worse (2), the same (3), better (4) or much better (5) than the study GPA of peers at the same higher education institution in the same subject of studies and semester. Independent variables: see text.

4.7 Conclusion

Two thirds of all German merit-based aid holders come from families where at least one parent achieved a college degree, whereas students of academic background only make up half of the overall student population (Middendorff et al., 2009, p. 24). Middendorff et al. (2009) argue that the likelihood to encounter students of non-academic backgrounds in the group of qualified students is lower than the likelihood to come across students whose parents have studied. Studies on differences in academic abilities between students who have already made their way to higher education are, however, not available for Germany. Yet, international evidence suggests that gaps in college grades are very small (Delaney et al., 2011; Aspelmeier et al., 2012) and not comparable to socio-economic differences in grades at earlier stages.

As the German merit-based aid system is very intransparent and the selection process of new scholarship holders is complex, it seems more reasonable that information asymmetries contribute to explaining the underrepresentation of non-academic students. This chapter is the first to investigate whether non-academic students qualified to receive merit-based scholarships apply indeed as frequently and are equally well informed about scholarship opportunities as are similar students of academic homes. If qualified students of college inexperienced families apply less often, although they might profit most from the scholarships' advantages, the merit-based system cannot unfold non-academic students' talent, thereby allocating funds inefficiently, and undermining its social mandate.

The findings from this chapter provide first evidence that participants in a field experiment were indeed significantly less informed at baseline if descending from families without academic experience. Keeping educational achievements, cognitive test scores, important application requirements, and a range of other covariates constant, students of non-academic backgrounds were also significantly less likely to report former applications for merit-based aid. Therefore, even if students of all socio-economic groups are equally likely to succeed in the application process for scholarships, the smaller share of non-academic students' applications will carry over to their underrepresentation in the scholarship body.

Nevertheless, if lower application rates are mainly resulting from information asymmetries, providing information about scholarship opportunities is a very inexpensive instrument to influence students' choice sets after leaving high school. The findings here suggest that providing information on scholarships increased non-academic students' knowledge on scholarships and led them to consider applying. Moreover, factual application rates of non-academic students six months later doubled with respect to the control group when a scholar with similar characteristics shared custom-fit information. As not all scholarship providers' application deadlines fell into the time span of six months, it is very likely that the treatment effects would be even larger if applications were questioned 9 or 12 months later. Moreover, these results represent a lower-bound estimate of the effect as the sample contained a majority of students formally ineligible to apply for funding. Restricting the sample to the highly eligible students increases the intent-to-treat effects for non-academic students in the role model treatment group substantially.

At the same time, providing publicly available information alone increased the awareness of scholarships in general and triggered applications for other, less selective ones. Yet, general information was not suitable to affect applications for highly selective merit-based aid. This finding is in line with previous evidence from the information interventions literature, suggesting that providing general information exerts no behavioral outcomes in industrialized countries (Booij et al., 2012; Carrell and Sacerdote, 2013; Kerr et al., 2014). Therefore, the decisive information asymmetry is not the ignorance of mere facts about scholarships, but rather the information that a similar person made it.

I find no differences in the baseline probabilities to succeed between applicants of different socio-economic backgrounds. As it is a declared goal to increase the number of scholars from underrepresented groups, the higher number of qualified students' applications is very likely to translates into a higher number of scholarship winners.

My results do, however, also suggest that female participants' applications were unaffected after offered detailed information, while men seem to have embraced the opportunity to apply. The findings from this chapter provide evidence that women underestimate their own abilities with respect to their peers—may it be because they are generally less confident about their own performances or because the average level of performance in a female-dominated peer group is higher. As merit-based scholarships are awarded in a highly demanding selection process and the role model treatment provided detailed information on its competitiveness, gender differences in competitiveness might also explain why women in the second treatment group did not apply more often. A wide range of studies provide evidence that women shy away from competition, while men embrace it and even perform better when competing (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007; Morin, 2015). With respect to merit-based aid, Kuhlmann et al. (2012) provide evidence that women having been recommended to the largest German scholarship providing institution are less successful in the assessment centers than their male counterparts, although equally well qualified. Learning about details of the later selection process might, therefore, shift women's lower odds to succeed in the process to an earlier stage: Anticipating the challenge to compete and potential problems to prevail in the process, women might abstain from applying in the first place. More evidence is, however, needed to investigate reasons for the gender gap and assess whether the findings from this non-representative sample can be generalized to the full student population. Accordingly, prospective studies should include a direct measure of participants' tastes for competition and level of self-confidence to set limits to possible reasons of the gender gap.

Some BFW have already established small mentoring programs where current scholarship holders get in touch with students from underrepresented groups and share information on scholarships. These programs are highly cost-effective as scholarship holders act on a volunteer work basis. The results from this chapter suggest that these programs can indeed be a fruitful and inexpensive endeavor to promote nonacademic students.

4.8 Appendix

4.8.1 Robustness to matching quality

A matching algorithm allocated each member of the role model treatment group to the most similar role model. To draw from the pool of available role models, the algorithm matched political party identification and religious denomination in a first step. If several role models were available on that basis, a role model of the same field of studies and/or gender was randomly selected. For 1% of participants, the algorithm could not select a matching role model in the first step, e.g., if the participant indicated to be socially engaged in a religious denomination not covered by the German BFWs. In that case, only field of studies and/or gender were matched. If there was more than one most similar role model, the algorithm randomly allocated the participant to one role model. Due to this procedure, the level of similarity to the role model differed slightly between participants and might have introduced bias.

If a higher quality of matching positively impacted participants' application behavior, controls accounting for similarity to the role model should be significantly positive. Adding controls for all matching dimensions (dummy = 1 if characteristics coincide, 0 otherwise) in column 1 of table A4.11, I do not find any of the dummies statistically significantly different from zero. To explore whether matching quality might have been more important for students of non-academic backgrounds or women and could therefore account for significant treatment effects found, I interact these variables with similarity controls in columns 2 and 3. I do again not find statistically significant effects. I rerun these analyses in table A4.12 but sum up the total number of similarities. Taking participants who were matched on half of the matching criteria as a reference group, those matched worse should have been less and those matched better should have been more likely to apply if similarity had a positive and relevant impact. Again, no clear pattern with respect to signs of coefficients evolves and none of the dummies is statistically significantly different from zero. Additionally, both the differential effect for students of non-academic backgrounds and women are robust to the inclusion of similarity indicators. Potentially different matching qualities between different student groups can, therefore, not explain different application rates.

It is luring but false to conclude from this analysis that similarity to the role model did not matter at all. As I had to maximize similarity in order to secure the relevance of the provided information for the treated, the variation in matching quality between participants is rather small, thereby impeding the probability to detect significant effects. Moreover, not all information to

Table A4.11: Influence of similarity on applications in the role model treatment group

	(1)	(2)	(3)
Non-academic background	0.024***	−0.037	0.022***
	(0.008)	(0.069)	(0.008)
Female	−0.019*	−0.018*	−0.110
	(0.010)	(0.010)	(0.073)
Matching criteria			
Same party	−0.015	0.002	−0.045
	(0.016)	(0.019)	(0.035)
Same religious denomination	−0.033	−0.082	−0.047
	(0.034)	(0.052)	(0.055)
Same field of studies	0.006	0.014	0.020
	(0.009)	(0.012)	(0.017)
Same gender	−0.002	0.004	−0.021
	(0.010)	(0.011)	(0.028)
Interactions with matching criteria			
Same party * Non-academic		−0.031	
		(0.031)	
Same religious denom. * Non-academic		0.103	
		(0.065)	
Same field * Non-academic		−0.014	
		(0.017)	
Same gender * Non-academic		−0.010	
		(0.019)	
Same Party * Female			0.055
			(0.039)
Same religious denom. * Female			0.031
			(0.066)
Same field * Female			−0.024
			(0.020)
Same gender * Female			0.034
			(0.029)
Observations	1730	1730	1730

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: The table contains results of OLS regressions of the application in wave 2 for the second treatment group only. Results from non-linear models are similar. Each estimation controls for the covariates of table 4.1, including former applications at a BFW and the former receipt of other scholarships at baseline. All similarity dummies are equal to 1 if the characteristics of the participant and the role model coincide and 0 otherwise.

Table A4.12: Influence of similarity on applications in the role model treatment group by number of similarities

	(1)	(2)	(3)
Non-academic background	0.023***	0.028*	0.022***
	(0.008)	(0.017)	(0.008)
Female	−0.019*	−0.018*	−0.048
	(0.010)	(0.010)	(0.032)
Number of similarities			
One of Four	0.012	0.025	0.004
	(0.033)	(0.037)	(0.058)
Three of four	0.015	0.011	−0.013
	(0.011)	(0.013)	(0.034)
Four of four	−0.003	0.012	−0.035
	(0.010)	(0.014)	(0.033)
Interactions with no. of similarities			
One * Non-academic		−0.025	
		(0.068)	
Three * Non-academic		0.005	
		(0.021)	
Four * Non-academic		−0.027	
		(0.020)	
One * Female			−0.018
			(0.060)
Three * Female			0.034
			(0.036)
Four * Female			0.039
			(0.035)
Observations	1730	1730	1730

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: See notes of table A4.11.

assess the overall degree of similarity was collected for all participants. For example, participants were only asked about their religious denomination if socially engaged in church. Attachment to church might be most relevant for participants with volunteer work in church. Nevertheless, religious but socially not engaged participants could also feel close to a matched role model of a religious BFW, although I cannot control for this match.[26]

To explore whether overall fit between participant and matched BFW mattered, I regress participants' scholarship applications on a self-reported evaluation of personal fit with the BFW they were matched to.[27] Note that the self-assessed fit item asked respondents to evaluate the similarity to the BFW funding the role model rather than to the role model. Therefore, I cannot separate the effect of similarity between participant and the BFW's association from the effect of similarity between the participant and the specific role model. Table A4.13 reveals that a good or very good self-assessed fit increases the probability to have applied by highly statistically significant 3.6 and 3.7 percentage points.

A last issue addressed here is whether slight differences in content or writing style between interview texts might have influenced application rates significantly—apart from similarity to the role model. I regress application behavior on dummies for all 34 interview texts, taking the text which was most frequently drawn by the algorithm as the reference category, and controlling for the quality of matching (not reported, results available on request). I find only one of the 33 interview dummies statistically significantly different from zero on the 5%-level—which is in line with a usual rate of false discoveries in multiple testing. Moreover, this text was shown to less than 10% of participants in the role model treatment group and should, therefore, not affect the results.

---

26 Religious denomination was coded to be similar (=1) if participants reporting volunteer work within church were matched with a BFW of equal religious denomination. Religious denomination was coded to be dissimilar (=0) if matched with a BFW of other religious denomination. Participants without religious volunteer work were coded as 1 if matched with a non-religious BFW. This coding takes into account that religious BFWs favor applicants socially engaged in church and with the same religious denomination. I also tested an alternative coding setting the similarity dummy only for those participants to 1 who were matched according to their religious engagement, considering all others as unmatched. Although this coding introduces an imbalance between religious and not religious participants—the latter always considered to be matched worse even if they might perfectly identify with the matched non-religious role model—the similarity dummy stays statistically insignificant.

27 The self-assessed fit question in wave 2 was worded as follows: "Please think back to the last survey. You have read an interview with a < male/female > scholar of < name of the BFW >. If you wanted to apply for a scholarship, how good would this BFW fit your personal political, religious, and ideological attitude?"

Table A4.13: Influence of self-assessed fit on applications in the role model treatment group

	(1)	(2)	(3)
Non-academic background	0.031**	0.030**	0.030**
	(0.012)	(0.012)	(0.012)
Female	−0.023*	−0.020	−0.021
	(0.013)	(0.013)	(0.014)
Self-assessed personal fit with BFW			
(Very) good fit	0.036***	0.037***	0.037***
	(0.013)	(0.013)	(0.013)
(Very) bad fit	0.009	0.008	0.010
	(0.014)	(0.014)	(0.014)
Matching criteria			
Same party		−0.012	
		(0.020)	
Same religious denomination		−0.040	
		(0.049)	
Same field of studies		0.013	
		(0.012)	
Same gender		0.007	
		(0.013)	
Number of similarities			
One of four			−0.009
			(0.040)
Three of four			0.020
			(0.016)
Four of four			0.003
			(0.016)
Observations	1110	1110	1110

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes: The table contains results of OLS regressions of the application in wave 2 on a set of covariates for the second treatment group only. Results from non-linear models are similar. Each estimation controls for the covariates of table 4.1, including former applications at a BFW and the former receipt of other scholarships at baseline. Reference category of the self-assessed fit variable is "partly, partly" fit between the respondent and the matched BFW. I dropped those who answered "don't know" (approximately 12% of cases) and for whom self-assessed fit is, accordingly, missing.

4.8.2 Additional figures

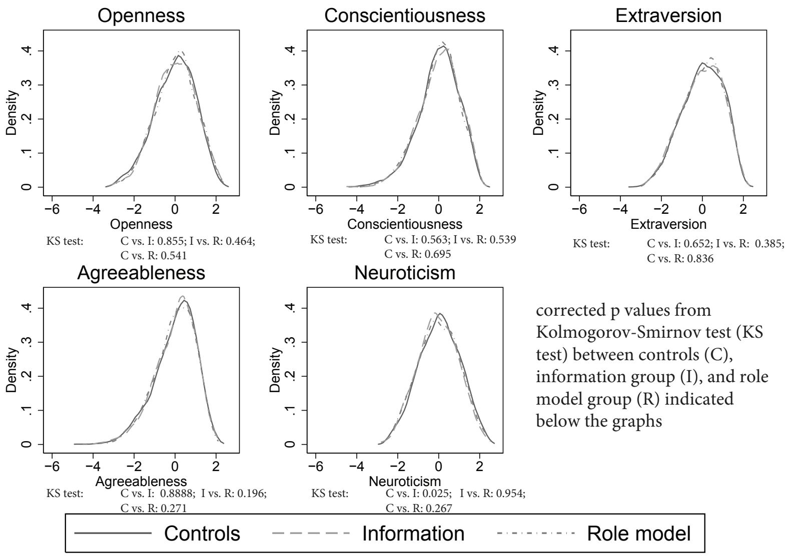


Figure A4.2: Differences in the Big Five Inventory between experimental groups

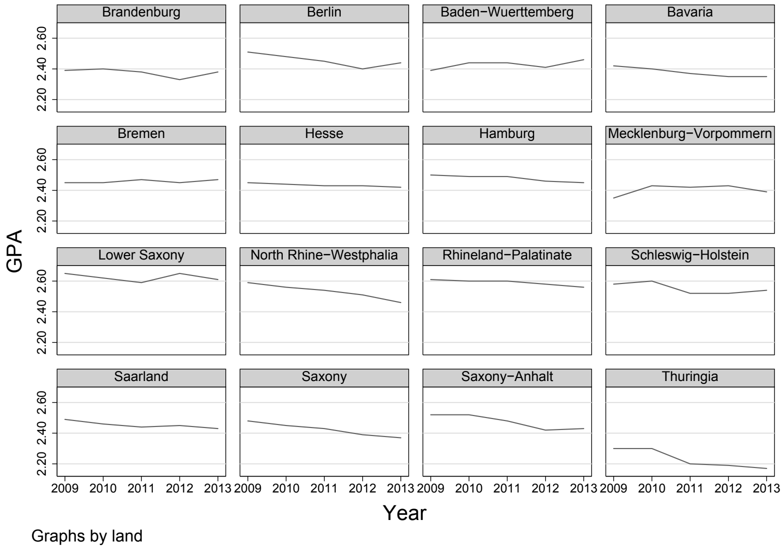


Figure A4.3: Grade inflation in German high school leaving certificates

Notes: Average GPAs of high school certificates in all 16 German states over time. Data source: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland IVC/Statistik (2015).

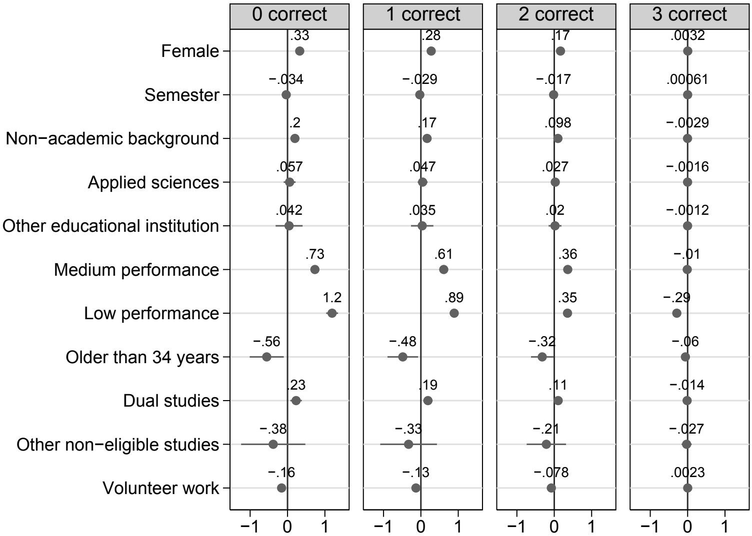


Figure A4.4: Information asymmetries over cut-offs (1/2)

Notes: Average semi-elasticities from an ordered logit model using 95%-confidence intervals. Positive (negative) values indicate a higher (lower) likelihood to answer the respective number of items correctly.

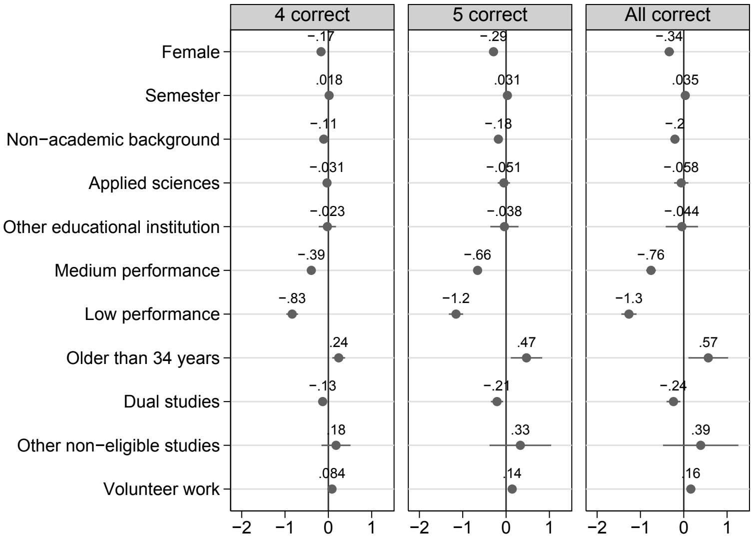


Figure A4.5: Information asymmetries over cut-offs (2/2)

Notes: See figure A4.4.

4.8.3 Additional tables

Table A4.14: Pre-treatment reasons for not applying

	Mean	(S.D.)
My grades are not good enough.	3.03	(0.90)
My voluntary work is not sufficient.	3.05	(0.91)
The funding amount I would receive is too small to be worth the application.	1.62	(0.76)
I know too little about the application requirements.	3.18	(0.86)
I received too little support by my college lecturers.	2.47	(1.03)
I do not need a scholarship as I can draw on other financial sources.	2.49	(0.94)
The application process is too complicated.	2.58	(0.90)
I do not want to incur liabilities tied to funding, e.g., seminar participation.	2.30	(0.97)
Observations	2670	

Notes: The sample size is smaller as only participants who mentioned to never have applied for a scholarship at wave 1 were questioned. The exact wording of the question was: "The following list contains reasons for why some students do not apply for a scholarship. In how far do these reasons also apply to you?" Participants rated each answer on a 4-point scale from "1 – Does not apply at all" to "4 – Applies fully". The order of the items in the table equals the order in which they were asked in the survey.

Table A4.15: Application requirements of the German Begabtenförderungswerke with religious association

Association	Name	Eligible semesters	Eligibility criteria	Formal requirements	Application deadline
Muslim	Avicenna-Studienwerk	BA: 1-2; MA: before start	High performance; social engagement; confession to Islam; interest in intercultural dialogue	No part-time studies; Muslim confession—exceptions possible on rare occasions	01.04.; 01.10. (still unknown at the time of the experiment because newly established)
Catholic	Cusanuswerk	BA: 1-2; MA: before start	High performance; interdisciplinary focus; personality; social engagement; attachment to church	Catholic confession	01.07.
Jewish	Ernst-Ludwig-Ehrlich Studienwerk	BA: 1; MA: before start	High performance; social engagement in Jewish community or other areas	Attachment to Jewish community	01.07.; 01.01.
Protestant	Evangelisches Studienwerk Villigst	BA: 1-2; no MA	Social engagement; enthusiasm in own studies; good performance; interest in interdisciplinary topics	Attachment to protestant church favored; exceptions for students > 35 or in second degrees possible; no part-time, distance, or dual studies	01.03.; 01.09.

Notes: Wording and order of different eligibility criteria and formal requirements follow the current information provided on the respective BFWs' websites. Applicants' number of semesters may not exceed the number of maximal semesters (see "Eligible semesters") when applying—calculations were based on a prescribed period of study of 6 semesters in the Bachelor's (BA) and 4 semesters in the Master's (MA) degree. Programs for PhDs and special programs for certain fields of studies etc. are not contained. Abbreviations: STEM = science, technology, engineering, and mathematics; uni AS = university of applied sciences.

Table A4.16: Application requirements of the German Begabtenförderungswerke with political association

Association	Name	Eligible semesters	Eligibility criteria	Formal requirements	Application deadline
Green	Heinrich Boell Foundation	BA: 1-3; no MA	Outstanding performance; interest in social/political issues; personality; supports the foundation's goals; focus on non-traditional, female, STEM-, and uni AS students	No part-time, dual degree, or second degree studies; no age restriction	Uni: 15.1., 15.7.; uni AS: 31.05., 30.11.
Christian, liberally conservative	Konrad Adenauer Foundation	BA: 1-2 if no MA intended; otherwise: 1-6; MA: before start	Cognitive abilities; high performance; interdisciplinary interest; supports the foundation's values; social engagement; personality	Dual studies possible if not part-time; no second degrees; distance studies possible; <36 years	15.01.; 01.07.
Democratic socialism	Rosa Luxemburg Foundation	BA: 2-3; MA: 1	High performance; social engagement in cohesion with socialist/foundation's values; female, needy, and non-academic students favored	Dual studies possible if not part-time; no second degrees	15.10.; 15.04.

Table A4.16: Continued

Association	Name	Eligible semesters	Eligibility criteria	Formal requirements	Application deadline
Green	Heinrich Boell Foundation	BA: 1-3; no MA	Outstanding performance; interest in social/political issues; personality; supports the foundation's goals; focus on non-traditional, female, STEM-, and uni AS students	No part-time, dual degree, or second degree studies; no age restriction	Uni: 15.1., 15.7.; uni AS: 31.05., 30.11.
Christian, liberally conservative	Konrad Adenauer Foundation	BA: 1-2 if no MA intended; otherwise: 1-6; MA: before start	Cognitive abilities; high performance; interdisciplinary interest; supports the foundation's values; social engagement; personality	Dual studies possible if not part-time; no second degrees; distance studies possible; < 36 years	15.01.; 01.07.
Democratic socialism	Rosa Luxemburg Foundation	BA: 2-3; MA: 1	High performance; social engagement in cohesion with socialist/foundation's values; female, needy, and non-academic students favored	Dual studies possible if not part-time; no second degrees	15.10.; 15.04.

Notes: See table A4.15. * Special requirements with respect to financial need, subject, and high school GPA for freshmen.

Table A4.17: Application requirements of the German Begabtenförderungswerke with other association

Association	Name	Eligible semesters	Eligibility criteria	Formal Requirements	Application deadline
Company-close	Foundation of German Business	BA: 1-2 if no MA intended; otherwise: BA: 1-6; MA: before start	High performance; social engagement; single-mindedness; social competence; general knowledge; joined-up thinking; ability to communicate	No second degree students; $<$ 33 years	Flexible, depends on location
Ideologically neutral	German National Scholarship Foundation	Proposal: BA: 2-4, MA: 1-2; application: BA: 1-2, no MA	"Performance, initiative, responsibility"	No second degree students; no part-time students; dual studies possible; $<$ 35 years	18.02.
Close to trade unions	Hans Boeckler Foundation	Flexible	Financial need; high achieving; social engagement	Special requirements for different programs, e.g., membership in a trade union	Depends on type of program

Notes: See table A4.15.

Bibliography

Alesina, Alberto and Nicola Fuchs-Schündeln (2007). Good-Bye Lenin (or Not?): The Effect of Communism on People's Preferences, *American Economic Review*, **97** (4), pp. 1507–1528.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström (2008). Eliciting Risk and Time Preferences, *Econometrica*, **76** (3), pp. 583–618.

Anderson, Patricia M. and Bruce D. Meyer (1997). Unemployment Insurance Takeup Rates and the After-Tax Value of Benefits, *The Quarterly Journal of Economics*, **112** (3), pp. 913–937.

Aries, Myriam B.C. and Guy R. Newsham (2008). Effect of Daylight Saving Time on Lighting Energy Use: A Literature Review, *Energy Policy*, **36** (6), pp. 1858–1866.

Arneklev, Bruce J., Lori Elis, and Sandra Medlicott (2006). Testing the General Theory of Crime: Comparing the Effects of "Imprudent Behavior" and an Attitudinal Indicator of "Low Self-Control", *Western Criminology Review*, **7** (3), pp. 41–55.

Aspelmeier, Jeffery E., Ann N. Elliott, Michael M. Love, Lauren A. McGill, and Thomas W. Pierce (2012). Self-Esteem, Locus of Control, College Adjustment, and GPA Among First- and Continuing-Generation Students: A Moderator Model of Generational Status, *Research in Higher Education*, **53** (7), pp. 755–781.

Astill, Rebecca G., Van der Heijden, Kristiaan B., Van Ijzendoorn, Marinus H., and Van Someren, Eus J. W. (2012). Sleep, Cognition, and Behavioral Problems in School-Age Children: A Century of Research Meta-Analyzed, *Psychological Bulletin*, **138** (6), pp. 1109–1138.

Avdic, Daniel and Marie Gartell (2015). Working While Studying? Student Aid Design and Socioeconomic Achievement Disparities in Higher Education, *Labour Economics*, **33**, pp. 26–40.

Banks, Siobhan and David F. Dinges (2007). Behavioral and Physiological Consequences of Sleep Restriction, *Journal of Clinical Sleep Medicine*, **3** (5), pp. 519–528.

Bargain, Olivier, Herwig Immervoll, and Heikki Viitamäki (2012). No Claim, no Pain: Measuring the Non-Take-Up of Social Assistance Using Register Data, *The Journal of Economic Inequality*, **10** (3), pp. 375–395.

Barnes, Christopher M. and David T. Wagner (2009). Changing to Daylight Saving Time Cuts into Sleep and Increases Workplace Injuries, *The Journal of Applied Psychology*, **94** (5), pp. 1305–1317.

Baum, Sandy and Saul Schwartz (2006). *How Much Debt is Too Much? Defining Benchmarks for Manageable Student Debt*: College Board.

Baumeister, Roy F., Kathleen D. Vohs, and Dianne M. Tice (2007). The Strength Model of Self-Control, *Current Directions in Psychological Science*, **16** (6), pp. 351–355.

Baumgartner, Hans J. and Viktor Steiner (2005). Student Aid, Repayment Obligations and Enrolment in Higher Education in Germany – Evidence from a "Natural Experiment", *Schmollers Jahrbuch* (125), pp. 29–38.

Baumgartner, H.J and V. Steiner (2006). Does More Generous Student Aid Increase Enrolment Rates into Higher Education? Evaluating the German Student Aid Reform of 2001, *IZA Discussion Paper* (2034).

Becker, Gary S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago: The University of Chicago Press, 3rd edition.

Becker, Irene and Richard Hauser (2003). Nicht-Inanspruchnahme zustehender Sozialhilfeleistungen (Dunkelzifferstudie): Endbericht zur Studie im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung.

Behncke, Stefanie, Markus Frölich, and Michael Lechner (2010). A Caseworker Like Me: Does The Similarity Between The Unemployed and Their Caseworkers Increase Job Placements? *The Economic Journal*, **120** (549), pp. 1430–1459.

Belley, Philippe and Lance Lochner (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement, *Journal of Human Capital*, **1** (1), pp. 37–89.

Berns, Gregory S., David Laibson, and George Loewenstein (2007). Intertemporal Choice: Toward an Integrative Framework, *Trends in Cognitive Sciences*, **11** (11), pp. 482–488.

Bettinger, Eric P. and Rachel Baker (2011). The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring, *NBER Working Paper* (16881).

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment, *The Quarterly Journal of Economics*, **127** (3), pp. 1205–1242.

BFW Working Group (2013). Stipendiumplus: Begabtenförderung im Hochschulbereich, URL: http://www.stipendiumplus.de, accessed on 10.10.2013.

Bird, Kelli and Benjamin L. Castleman (2014). Here Today, Gone Tomorrow? Investigating Rates and Patterns of Financial Aid Renewal Among College Freshmen, *EdPolicyWorks Working Paper Series* (25).

Björkstén, Karin S., Daniel F. Kripke, and Peter Bjerregaard (2009). Accentuation of Suicides but not Homicides with Rising Latitudes of Greenland in the Sunny Months, *BMC Psychiatry*, **9**, p. 20.

Black, Sandra E. and Paul J. Devereux (2011). Recent Developments in Intergenerational Mobility, in David Card and Orley Ashenfelter (Ed.) *Handbook of Labor Economics*, **Volume 4, Part B**: Elsevier, pp. 1487–1541.

Blossfeld, Pia N., Gwendolin J. Blossfeld, and Hans-Peter Blossfeld (2016). Changes in Educational Inequality in Cross-National Perspective, in J. Michael Shanahan, T. Jeylan Mortimer, and Monica Kirkpatrick Johnson (Eds.) *Handbook of the Life Course: Volume II*, Cham: Springer International Publishing, pp. 223–247.

Blundell, Richard, Vanessa Fry, and Ian Walker (1988). Modelling the Take-up of Means-Tested Benefits: The Case of Housing Benefits in the United Kingdom, *The Economic Journal*, **98** (390), pp. 58–74.

Boatman, Angela, Brent Evans, and Adela Soliz (2014). Applying the Lessons of Behavioral Economics to Improve the Federal Student Loan Programs: Six Policy Recommendations, *Lumina Foundation Report*, **April**.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek (2012). The Role of Information in the Take-Up of Student Loans, *Economics of Education Review*, **31** (1), pp. 33–44.

Borghans, Lex, Bart H. H. Golsteyn, Anders Stenberg, and Jacobus van Wouwe (2015). Does Expert Advice Improve Educational Choice?, *PLOS ONE*, **10** (12), p. e0145378.

Bors, D. A. and T. L. Stokes (1998). Raven's Advanced Progressive Matrices: Norms for First-Year University Students and the Development of a Short Form, *Educational and Psychological Measurement*, **58** (3), pp. 382–398.

Brosig-Koch, Jeannette, Christoph Helbach, Axel Ockenfels, and Joachim Weimann (2011). Still Different After All These Years: Solidarity Behavior in East and West Germany, *Journal of Public Economics*, **95** (11-12), pp. 1373–1376.

Bruckmeier, Kerstin, Johannes Pauser, Ulrich Walwei, and Jürgen Wiemers (2013). Simulationsrechnungen zum Ausmaß der Nicht-Inanspruchnahme von Leistungen der Grundsicherung: Studie im Auftrag des Bundesministeriums für Arbeit und Soziales zur Abgrenzung und Struktur von Referenzgruppen für die Ermittlung von Regelbedarfen auf Basis der Einkommens- und Verbrauchsstichprobe 2008, URL: http://EconPapers.repec.org/RePEc:iab:iabfob:201305.

Bruckmeier, Kerstin and Jürgen Wiemers (2012). A New Targeting: A New Take-Up? *Empirical Economics*, **43** (2), pp. 565–580.

Bruckmeier, Kerstin and Berthold U. Wigger (2014). The Effects of Tuition Fees on Transition from High School to University in Germany, *Economics of Education Review*, **41**, pp. 14–23.

Bucher-Koenen, Tabea and Bettina Lamla (2014). The Long Shadow of Socialism: On East-West German Differences in Financial Literacy, *Max Planck Institute for Social Law and Social Policy Discussion Paper* (282-14).

Bundeskanzleramt and Nationaler Normenkontrollrat (2010). Einfacher zum Studierenden-BAföG: Abschlussbericht, URL: http://bit.ly/VVOkzs.

Burgess, Helen J., Carlo S. Legasto, Louis F. Fogg, and Mark R. Smith (2013). Can Small Shifts in Circadian Phase Affect Performance? *Applied Ergonomics*, **44** (1), pp. 109–111.

Cadena, Brian C. (2008). How Immigrants and Students Respond to Public Policies: Evidence from Welfare Reform, the Minimum Wage and Stafford Loans, Ph.D. dissertation, The University of Michigan, Michigan.

Cadena, Brian C. and Benjamin J. Keys (2013). Can Self-Control Explain Avoiding Free Money? Evidence from Interest-Free Student Loans, *Review of Economics and Statistics*, **95** (4), pp. 1117–1129.

Callender, Claire (2008). The Impact of Term–Time Employment on Higher Education Students' Academic Attainment and Achievement, *Journal of Education Policy*, **23** (4), pp. 359–377.

——— (2009). *Awareness, Take-up and Impact of Institutional Bursaries and Scholarships in England: Summary and Recommendations*, Bristol: Office for Fair Access, URL: http://tinyurl.com/Office4FairAccess, accessed on 13.10.2014.

Cameron, Stephen V. and James J. Heckman (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males, *Journal of Political Economy*, **106** (2), pp. 262–333.

——— (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males, *Journal of Political Economy*, **109** (3), pp. 455–499.

Campbell, I. G., K. J. Grimm, E. de Bie, and I. Feinberg (2012). Sex, Puberty, and the Timing of Sleep EEG Measured Adolescent Brain Maturation, *Proceedings of the National Academy of Sciences*, **109** (15), pp. 5740–5743.

Carneiro, Pedro and James J. Heckman (2002). The Evidence on Credit Constraints in Post-Secondary Schooling, *The Economic Journal*, **112** (482), pp. 705–734.

Carrell, Scott E., Teny Maghakian, and James E. West (2011). A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents, *American Economic Journal: Economic Policy*, **3** (3), pp. 62–81.

Carrell, Scott E. and Bruce Sacerdote (2013). Late Interventions Matter Too: The Case of College Coaching New Hampshire, *NBER Working Paper* (19031).

Carskadon, Mary A., Cecilia Vieira, and Christine Acebo (1993). Association Between Puberty and Delayed Phase Preference, *Sleep: Journal of Sleep Research & Sleep Medicine*, **16** (3), pp. 258–262.

Castleman, Benjamin L., Lindsay C. Page, and Korynn Schooley (2014). The Forgotten Summer: Does the Offer of College Counseling After High School Mitigate Summer Melt Among College-Intending, Low-Income High School Graduates? *Journal of Policy Analysis and Management*, **33** (2), pp. 320–344.

Central Intelligence Agency (2013). The World Factbook 2013-14, URL: http://tinyurl.com/factb.

Chevalier, Arnaud, Steve Gibbons, Andy Thorpe, Martin Snell, and Sherria Hoskins (2009). Students' Academic Self-Perception, *Economics of Education Review*, **28** (6), pp. 716–727.

Cho, Soo Hyun, Yilan Xu, and D. Elizabeth Kiss (2015). Understanding Student Loan Decisions: A Literature Review, *Family and Consumer Sciences Research Journal*, **43** (3), pp. 229–243.

Coady, D., C. Martinelli, and S. W. Parker (2013). Information and Participation in Social Programs, *The World Bank Economic Review*, **27** (1), pp. 149–170.

College Board (Ed.) (2014). *SAT One-Year Mean Score Changes*, New York, URL: http://tinyurl.com/CollBoard.

Currie, Janet (2004). The Take Up of Social Benefits, *National Bureau of Economic Research Working Paper Series*, **10488**.

Deaux, Kay and Elizabeth Farris (1977). Attributing Causes for One's Own Performance: The Effects of Sex, Norms, and Outcome, *Journal of Research in Personality*, **11** (1), pp. 59–72.

Delaney, Liam, Colm Harmon, and Cathy Redmond (2011). Parental Education, Grade Attainment and Earnings Expectations Among University Students, *Economics of Education Review*, **30** (6), pp. 1136–1152.

DESTATIS, WZB, and SOEP (Eds.) (2013). *Datenreport 2013: Ein Sozialbericht für die Bundesrepublik Deutschland*, Bonn: Bundeszentrale für politische Bildung.

Dinkelman, Taryn and Claudia Martínez A. (2014). Investing in Schooling in Chile: The Role of Information About Financial Aid for Higher Education, *The Review of Economics and Statistics*, **96** (2), pp. 244–257.

Dorsett, Richard and Christopher Heady (1991). The Take-Up of Means-Tested Benefits by Working Families with Children, *Fiscal Studies*, **12** (4), pp. 22–32.

Draper, Electra (05.03.2015). Daylight Saving Time Change Impacts Schools, ERs and Roadways, *The Denver Post*, **2015**, URL: http://tinyurl.com/DSTschools.

Duckworth, Angela Lee and Margaret L. Kern (2011). A Meta-Analysis of the Convergent Validity of Self-Control Measures, *Journal of Research in Personality*, **45** (3), pp. 259–268.

Dynarski, Susan (2002). The Behavioral and Distributional Implications of Aid for College, *The American Economic Review*, **92** (2), pp. 279–285.

——— (2004). The New Merit Aid, in Caroline M. Hoxby (Ed.) *College Choices*, Chicago: Univ. of Chicago Press, pp. 63–100, URL: http://www.nber.org/chapters/c10098.

Dynarski, Susan M. and Judith E. Scott-Clayton (2006). The Cost of Complexity in Federal Student Aid: Lessons from Optimal Tax Theory and Behavioral Economics, *National Tax Journal*, **59** (2), pp. 319–356.

Dynarski, Susan and Mark Wiederspan (2012). Student Aid Simplification: Looking Back and Looking Ahead, *National Tax Journal*, **65** (1), pp. 211–234.

Edwards, Finley (2012). Early to Rise? The Effect of Daily Start Times on Academic Performance, *Economics of Education Review*, **31** (6), pp. 970–983.

Federal Institute for Research on Building, Urban Affairs and Spatial Development (2013). *Wohnungsengpässe und Mietensteigerungen: Aktuelle Mietenentwicklungen in den Städten und Regionen*, **7**, BBSR-Analysen kompakt, Bonn: BBR.

Federal Ministry of Education and Research (2014a). Impulse für die Gesellschaft: Die Begabtenförderungswerke., URL: http://www.bmbf.de/de/294.php, accessed on 04.08.2014.

——— (2014b). *Zusätzliche Nebenbestimmungen zur Förderung begabter Studierender sowie begabter Nachwuchswissenschaftlerinnen und -wissenschaftler*, URL: http://tinyurl.com/Stipendienrichtlinien, accessed on 29.03.2016.

——— (2015a). Stipendienentwicklung für Studierende in den Jahren 2005 bis 2014, URL: http://tinyurl.com/Stipendientrends, accessed on 29.03.2016.

——— (2015b). Übersicht über die Zuwendungen des BMBF an die Begabtenförderungswerke und die Zahl der Geförderten nach Jahren, URL: http://tinyurl.com/Stipendienbetrag, accessed on 29.03.2016.

Federal Statistical Office (Ed.) (2014a). *Bildung und Kultur: Studierende an Hochschulen Wintersemester 2013/2014*, Wiesbaden.

Federal Statistical Office (2014b). Long Term Series of Student Numbers from 1975, URL: http://tinyurl.com/studnumb, accessed on 29.03.2016.

Federal Statistical Office (Ed.) (2014c). *Studierende an Hochschulen Vorbericht - Fachserie 11 Reihe 4.1 - Wintersemester 2013/2014*, Wiesbaden.

——— (Ed.) (2015a). *Ausbildungsförderung nach dem Bundesausbildungsförderungsgesetz (BAföG) 2014*, Fachserie 11 Reihe 7, Wiesbaden.

Federal Statistical Office (2015b). Long Term Series of Student Numbers from 1975, URL: http://tinyurl.com/studnumb75, accessed on 29.03.2016.

Ferguson, S. A., D. F. Preusser, A. K. Lund, P. L. Zador, and R. G. Ulmer (1995). Daylight Saving Time and Motor Vehicle Crashes: The Reduction in Pedestrian and Vehicle Occupant Fatalities, *American Journal of Public Health*, **85** (1), pp. 92–95.

Field, Erica (2009). Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School, *American Economic Journal: Applied Economics*, **1** (1), pp. 1–21.

Finlay, Keith, Leandro M. Magnusson, and Mark E. Schaffer (2013). Weakiv: Weak-Instrument-Robust Tests and Confidence Intervals for Instrumental-Variable (IV) Estimation of Linear, Probit and Tobit Models, URL: http://ideas.repec.org/c/boc/bocode/s457684.html.

Finn, Dan and Jo Goodship (2014). Take-Up of Benefits and Poverty: An Evidence and Policy Review, *Centre for Economic & Social Inclusion Report*, URL: http://tinyurl.com/FinnGoodship.

forsa Gesellschaft für Sozialforschung und statistische Analysen mbH (2015). Meinungen zur Zeitumstellung, URL: http://tinyurl.com/forsaDST, accessed on 14.04.2015.

Frick, Joachim R. and Olaf Groh-Samberg (2007). To Claim or Not to Claim: Estimating Non-Take-Up of Social Assistance in Germany and the Role of Measurement Error, *SOEPpapers on Multidisciplinary Panel Data Research* (53).

Gaski, John F. and Jeff Sagarin (2011). Detrimental Effects of Daylight-Saving Time on SAT Scores, *Journal of Neuroscience, Psychology, and Economics*, **4** (1), pp. 44–53.

Gau, Shur-Fen and Wei-Tsuen Soong (2003). The Transition of Sleep-Wake Patterns in Early Adolescence, *Sleep*, **26** (4), pp. 449–454.

Gerlitz, Jean-Yves and Jürgen Schupp (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP: Dokumentation der Instrumententwicklung BFI-S auf Basis des SOEP-Pretests 2005, *DIW Berlin Research Notes* (4).

German Bundestag (2010). Unterrichtung durch die Bundesregierung: Achtzehnter Bericht nach § 35 des Bundesausbildungsförderungsgesetzes zur Überprüfung der Bedarfssätze, Freibeträge sowie Vomhundertsätze und Höchstbeträge nach § 21 Absatz 2, *German Bundestag Printed Matter* (17/485 (19.01.2010)).

——— (2014). Unterrichtung durch die Bundesregierung: Zwanzigster Bericht nach § 35 des Bundesausbildungsförderungsgesetzes zur Überprüfung der Bedarfssätze, Freibeträge sowie Vomhundertsätze und Höchstbeträge nach § 21 Absatz 2, *German Bundestag Printed Matter* (18/460 (04.02.2014)).

——— (2015). Unterrichtung durch die Bundesregierung: Bericht über die Höhe des steuerfrei zu stellenden Existenzminimums von Erwachsenen und Kindern für das Jahr 2016 (10. Existenzminimumbericht), *German Bundestag Printed Matter* (18/3893 (30.01.2015)).

German National Scholarship Foundation (2014). Jahresbericht 2013, URL: http://tinyurl.com/scholfound, accessed on 29.03.2016.

Gibson, Matthew and Jeffrey Shrader (2014). Time Use and Productivity: The Wage Returns to Sleep, *University of California at San Diego, Economics Working Paper Series.*

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in Competitive Environments: Gender Differences, *The Quarterly Journal of Economics*, **118** (3), pp. 1049–1074.

Gneezy, Uri and Aldo Rustichini (2004). Gender and Competition at a Young Age, *The American Economic Review*, **94** (2), pp. 377–381.

Goel, Namni, Hengyi Rao, Jeffrey S. Durmer, and David F. Dinges (2009). Neurocognitive Consequences of Sleep Deprivation, *Seminars in Neurology*, **29** (4), pp. 320–339.

Gottfredson, Michael R. and Travis Hirschi (1990). *A General Theory of Crime*, Stanford and California: Stanford University Press.

Grave, Barbara S. and Mathias Sinning (2014). Why Don't We Just Give Them the Money? Financing Living Expenses of Students in Germany, in Bruce Chapman, Timothy Higgins, and Joseph E. Stiglitz (Eds.) *Income Contingent Loans: Theory, Practice and Prospects*: Palgrave Macmillan, pp. 109–124.

Gregory-Allen, Russell, Ben Jacobsen, and Wessel Marquering (2010). The Daylight Saving Time Anomaly in Stock Returns: Fact or Fiction? *Journal of Financial Research*, **33** (4), pp. 403–427.

Guérin, Nicole, Suzanne Boulenguiez, Alain Reinberg, Geneviéve Di Costanzo, Philippe Guran, and Yvan Touitou (1993). Weekly Changes in Psychophysiological Variables of 8- to 10-Year-Old School Girls, *Chronobiology International*, **10** (6), pp. 471–479.

Hahn, Elisabeth, Juliana Gottschling, and Frank M. Spinath (2012). Short Measurements of Personality: Validity and Reliability of the GSOEP Big Five Inventory (BFI-S), *Journal of Research in Personality*, **46** (3), pp. 355–359.

Hanushek, Eric A. (1999). The Evidence on Class Size, in Susan E. Mayer and Paul E. Peterson (Eds.) *Earning and Learning*, Washington and D.C and New York and N.Y: Brookings Institution Press and Russell Sage Foundation, pp. 131–168.

——— (2006). School Resources, in Eric A. Hanushek and F. Welch (Eds.) *Handbook of the Economics of Education*, **2**: Elsevier, pp. 865–908.

——— (2009). The Economic Value of Education and Cognitive Skills, in Gary Sykes, Barbara L. Schneider, David Nathan Plank, and Timothy G. Ford (Eds.) *Handbook of Education Policy Research*, New York and Washington D.C.: Routledge and American Educational Research Association, pp. 39–56.

Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutström (2005). Temporal Stability of Estimates of Risk Aversion, *Applied Financial Economics Letters*, **1** (1), pp. 31–35.

Hastings, Justine, Christopher Neilson, and Seth Zimmerman (2015). The Effects of Earnings Disclosure on College Enrollment Decisions, *NBER Working Paper* (21300).

Hastings, Justine S. and Jeffrey M. Weinstein (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments, *The Quarterly Journal of Economics*, **123** (4), pp. 1373–1414.

Hauty, G. T. and T. Adams (1965a). Phase Shifts of the Human Circadian System and Performance Deficit During the Periods of Transition. I. East-West Flight, *Office of Aviation Medicine Report, (AM 65-28).*

——— (1965b). Phase Shifts of the Human Circadian System and Performance Deficit During the Periods of Transition. II. West-East Flight, *Office of Aviation Medicine Report, (AM 65-29).*

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, in Eric A. Hanushek and F. Welch (Eds.) *Handbook of the Economics of Education*, **Volume 1**: Elsevier, pp. 307–458.

Heineck, Guido and Regina T. Riphahn (2009). Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades, *Jahrbücher für Nationalökonomie und Statistik*, **229** (1), pp. 36–60.

Heineck, Guido and Bernd Süssmuth (2013). A Different Look at Lenin's Legacy: Social Capital and Risk Taking in the Two Germanies, *Journal of Comparative Economics*, **41** (3), pp. 789–803.

Heissel, Jennifer and Samuel Norris (2015). Rise and Shine: The Effect of School Start Times on Academic Performance from Childhood Through Puberty, *Working Paper.*

Herber, Stefanie P. (2015). The Role of Information in the Application for Merit-Based Scholarships: Evidence From a Randomized Field Experiment, *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Education 3* (C07-V2), URL: http://hdl.handle.net/10419/113058.

Hernandez, Monica and Stephen Pudney (2007). Measurement Error in Models of Welfare Participation, *Journal of Public Economics*, **91** (1–2), pp. 327–341.

Hernanz, Virginia, Franck Malherbet, and Michele Pellizzari (2004). OECD Social, Employment and Migration Working Papers, *OECD Social, Employment and Migration Working Papers*, **17**.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina (2008). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends, *The B.E. Journal of Economic Analysis & Policy*, **7** (2).

Heschong, Lisa, Roger L. Wright, and Stacia Okura (2002). Daylighting Impacts on Human Performance in School, *Journal of the Illuminating Engineering Society*, **31** (2), pp. 101–114.

Hicks, R. A., K. Lindseth, and J. Hawkins (1983). Daylight Saving-Time Changes Increase Traffic Accidents, *Perceptual and Motor Skills*, **56** (1), pp. 64–66.

Hinrichs, Peter (2011). When the Bell Tolls: The Effects of School Starting Times on Academic Achievement, *Education Finance and Policy*, **6** (4), pp. 486–507.

Hochschulrektorenkonferenz (Ed.) (2012). *Statistische Daten zu Studienangeboten an Hochschulen in Deutschland: Studiengänge, Studierende, Absolventen Wintersemester 2012/2013*, Statistiken zur Hochschulpolitik, Bonn.

Hoxby, Caroline M. and Sarah Turner (2013). Informing Students about Their College Options: A Proposal for Broadening the Expanding College Opportunities Project, *The Hamilton Project Discussion Paper* (2013-03).

Hübner, Malte (2012). Do Tuition Fees Affect Enrollment Behavior? Evidence from a 'Natural Experiment' in Germany, *Economics of Education Review*, **31** (6), pp. 949–960.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, **47** (1), pp. 5–86.

Janszky, Imre, Staffan Ahnve, Rickard Ljung, Kenneth J. Mukamal, Shiva Gautam, Lars Wallentin, and Ulf Stenestrand (2012). Daylight Saving Time Shifts and Incidence of Acute Myocardial Infarction: Swedish Register of

Information and Knowledge About Swedish Heart Intensive Care Admissions (RIKS-HIA), *Sleep Medicine*, **13** (3), pp. 237–242.

Jensen, Robert (2010). The (Perceived) Returns to Education and the Demand for Schooling, *The Quarterly Journal of Economics*, **125** (2), pp. 515–548.

Joncas, Marc (2012). TIMSS 2011 Population Coverage and Exclusions, in Michael O. Martin and Ina V.S Mullis (Eds.) *Methods and Procedures in TIMSS and PIRLS 2011*, Chestnut Hill and MA: TIMSS & PIRLS International Study Center, Boston College.

Kamstra, Mark J., Lisa A. Kramer, and Maurice D. Levi (2000). Losing Sleep at the Market: The Daylight Saving Anomaly, *The American Economic Review*, **90** (4), pp. 1005–1011.

Kayser, Hilke and Joachim Frick (2001). Take It or Leave It: (Non-)Take-Up Behavior of Social Assistance in Germany, *Schmollers Jahrbuch* (1), pp. 27–58.

Kellogg, Ryan and Hendrik Wolff (2008). Daylight Time and Energy: Evidence from an Australian Experiment, *Journal of Environmental Economics and Management*, **56** (3), pp. 207–220.

Kerbusk, Simon (2009). Wer hat, dem wird gegeben: Studie zur Begabtenförderung, *DIE ZEIT* (N° 40), URL: http://www.zeit.de/2009/40/C-Begabtenfoerderung, accessed on 18.11.2014.

Kerr, Sari Pekkala, Tuomas Pekkarinen, Matti Sarvimäki, and Roope Uusitalo (2014). Educational Choice and Information on Labor Market Prospects: A Randomized Field Experiment, *Working Paper*.

Keys, Benjamin J. (2009). Three Essays on Labor and Credit Markets, Ph.D. dissertation, The University of Michigan, Michigan.

King, Jacqueline E. (2006). Missed Opportunities Revisited: New Information on Students Who Do Not Apply for Financial Aid, *American Council on Education Issue Brief*.

Kofoed, Michael S. (2015). To Apply or Not to Apply: FAFSA Completion and Financial Aid Gaps, *Working Paper*, URL: http://ssrn.com/abstract=2353846.

Kotchen, Matthew J. and Laura E. Grant (2011). Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana, *Review of Economics and Statistics*, **93** (4), pp. 1172–1185.

Kountouris, Yiannis and Kyriaki Remoundou (2014). About Time: Daylight Saving Time Transition and Individual Well-Being, *Economics Letters*, **122** (1), pp. 100–103.

Kuehnle, Daniel and Christoph Wunder (2015). Using the Life Satisfaction Approach to Value Daylight Savings Time Transitions: Evidence from Britain and Germany, *SOEPpapers (744)*.

Kuhlmann, Kerstin, Svenja Kremer, Jahid Hassan, and Georg Rudinger (2012). Evaluierung des Auswahlverfahrens der Studienstiftung des deutschen Volkes für die Studienstiftung des deutschen Volkes e. V., URL: http://www.studienstiftung.de/publikationen/abschlussbericht-zur-evaluierung-der-auswahlverfahren.html, accessed on 13.08.2014.

Laberge, L., D. Petit, C. Simard, F. Vitaro, R. E. Tremblay, and J. Montplaisir (2001). Development of Sleep Patterns in Early Adolescence, *Journal of Sleep Research*, **10** (1), pp. 59–67.

Lahti, T., J. Sysi-Aho, J. Haukka, and T. Partonen (2011). Work-Related Accidents and Daylight Saving Time in Finland, *Occupational Medicine (Oxford, England)*, **61** (1), pp. 26–28.

Laird, Donald A. (1925). Relative Performance of College Students as Conditioned by Time of Day and Day of Week, *Journal of Experimental Psychology*, **8** (1), pp. 50–63.

Lemmer, Björn, Ralph-Ingo Kern, Gudrun Nold, and Heinz Lohrer (2002). Jet Lag in Athletes after Eastward and Westward Time-Zone Transition, *Chronobiology International*, **19** (4), pp. 743–764.

Loyalka, Prashant, Yingquan Song, Jianguo Wei, Weiping Zhong, and Scott Rozelle (2013). Information, College Decisions and Financial Aid: Evidence From a Cluster-Randomized Controlled Trial in China, *Economics of Education Review*, **36**, pp. 26–40.

Marx, D. M. and J. S. Roman (2002). Female Role Models: Protecting Women's Math Test Performance, *Personality and Social Psychology Bulletin*, **28** (9), pp. 1183–1193.

Marx, David M. and Sei Jin Ko (2012). Superstars "Like" Me: The Effect of Role Model Similarity on Performance Under Threat, *European Journal of Social Psychology*, **42** (7), pp. 807–812.

McGarry, Kathleen (1996). Factors Determining Participation of the Elderly in Supplemental Security Income, *The Journal of Human Resources*, **31** (2), p. 331.

Meier, Stephan and Charles D. Sprenger (2015). Temporal Stability of Time Preferences, *The Review of Economics and Statistics*, **97** (2).

Middendorff, Elke, Beate Apolinarski, Jonas Poskowsky, Maren Kandulla, and Nicolai Netz (2013). *Die Wirtschaftliche und Soziale Lage der Studierenden in der Bundesrepublik Deutschland 2012: 20. Sozialerhebung des Deutschen Studentenwerks; Durchgeführt durch das HIS-Institut für Hochschulforschung*, Bonn u.a: Bundesministerium für Bildung und Forschung.

Middendorff, Elke, Wolfgang Isserstedt, and Maren Kandulla (2009). *Das soziale Profil in der Begabtenförderung: Ergebnisse einer Online-Befragung unter allen Geförderten der elf Begabtenförderungswerke im Oktober 2008*, HIS:Projektbericht, Hannover: HIS Hochschul-Informations-System GmbH.

Moffitt, Robert (1983). An Economic Model of Welfare Stigma, *The American Economic Review*, **73** (5), pp. 1023–1035.

Monk, Timothy H. and L. C. Aplin (1980). Spring and Autumn Daylight Saving Time Changes: Studies of Adjustment in Sleep Timings, Mood, and Efficiency, *Ergonomics*, **23** (2), pp. 167–178.

Morin, Louis-Philippe (2015). Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform, *The Journal of Labor Economics*, **33** (2), pp. 443–491.

Mullis, Ina V.S., Michael O. Martin, Ann M. Kennedy, Kathleen L. Trong, and Marian Sainsbury (2009a). *PIRLS 2011 Assessment Frameworks*, Lynch School of Education and Boston College: TIMSS & PIRLS International Study Center.

Mullis, Ina V.S., Michael O. Martin, Graham J. Ruddock, Christine Y. O'Sullivan, and Corinna Preuschoff (2009b). *TIMSS 2011 Assessment Frameworks*, Lynch School of Education and Boston College: TIMSS & PIRLS International Study Center.

National Center for Education Statistics (2013). Digest of Education Statistics, URL: http://tinyurl.com/NCESdigest.

Nguyen, Trang (2008). Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar, *MIT working paper, Cambridge: Massachusetts Institute of Technology.*

Niederle, Muriel and Lise Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, **122** (3), pp. 1067–1101.

Niu, Shu-Fen, Min-Huey Chung, Chiung-Hua Chen, Desley Hegney, Anthony O'Brien, and Kuei-Ru Chou (2011). The Effect of Shift Rotation on Employee Cortisol Profile, Sleep Quality, Fatigue, and Attention Level, *Journal of Nursing Research*, **19** (1), pp. 68–81.

O'Connor, Melissa C. and Sampo V. Paunonen (2007). Big Five Personality Predictors of Post-Secondary Academic Performance, *Personality and Individual Differences*, **43** (5), pp. 971–990.

OECD (2014). *Education at a Glance 2014: OECD Indicators*, Paris: OECD Publishing.

Oosterbeek, Hessel and Anja van den Broek (2009). An Empirical Analysis of Borrowing Behaviour of Higher Education Students in the Netherlands, *Economics of Education Review*, **28** (2), pp. 170–177.

Pabst, Josephine (05.04.2015). Stipendiaten auf Schatzsuche: Unbekannte Förderprogramme, URL: http://www.faz.net/-gyl-81j2f, accessed on 09.03.2016.

Page, Marianne E. (2004). New Evidence on the Intergenerational Correlation in Welfare Participation, in Miles Corak (Ed.) *Generational Income Mobility in North America and Europe*: Cambridge University Press, pp. 226–244.

Ramm, Michael (2014). Studierendensurvey und Repräsentativität: Dokumentation II, *Hefte zur Bildungs- und Hochschulforschung* (72).

Randler, Christoph (2011). Age and Gender Differences in Morningness–Eveningness During Adolescence, *The Journal of Genetic Psychology*, **172** (3), pp. 302–308.

Raven, John C., J. H. Court, and J. Raven (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 4)*, London: H.K. Lewis.

Reese, Margaret (1932). A Study of the Effect of Daylight Saving Time upon the Sleep of Young Children, *Child Development*, **3** (1), pp. 86–89.

Riphahn, Regina T. (2001). Rational Poverty or Poor Rationality? The Take-up of Social Assistance Benefits, *Review of Income and Wealth*, **47** (3), pp. 379–398.

Riphahn, Regina T. and Florian Schieferdecker (2010). The Transition to Tertiary Education and Parental Background over Time, *Journal of Population Economics*, **25** (2), pp. 635–675.

Rosen, Leora N., Steven D. Targum, Michael Terman, Michael J. Bryant, Howard Hoffman, Siegfried F. Kasper, Joelle R. Hamovit, John P. Docherty, Betty Welch, and Norman E. Rosenthal (1990). Prevalence of Seasonal Affective Disorder at Four Latitudes, *Psychiatry Research*, **31** (2), pp. 131–144.

Rothe, Friedrich and Ernst August Blanke (Eds.) (2015). *Bundesausbildungsförderungsgesetz: Kommentar*, **34**, 5th edition.

Rubin, Donals B. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley Series in Probability and Statistics, New York: Wiley.

Rubin, Rachel B. (2011). The Pell and the Poor: A Regression-Discontinuity Analysis of On-Time College Enrollment, *Research in Higher Education*, **52** (7), pp. 675–692.

Sandhu, Amneet, Milan Seth, and Hitinder S. Gurm (2014). Daylight Savings Time and Myocardial Infarction, *Open Heart*, **1** (e000019).

Saniter, Nils and Thomas Siedler (2014). The Effects of Occupational Knowledge: Job Information Centers, Educational Choices, and Labor Market Outcomes, *IZA Discussion Paper* (8100).

Schindler, Steffen and David Reimer (2010). Primäre und Sekundäre Effekte der Sozialen Herkunft beim Übergang in die Hochschulbildung, *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **62** (4), pp. 623–653.

Schmidt, Johannes (28.03.2009). Leistungsknick durch Zeitumstellung, *Frankfurter Rundschau*, **2009**, URL: http://tinyurl.com/Leistungsknick.

Schneider, Anne-Marie and Christoph Randler (2009). Daytime Sleepiness During Transition into Daylight Saving Time in Adolescents: Are Owls Higher at Risk? *Sleep Medicine*, **10** (9), pp. 1047–1050.

Schultz, Theodore W. (1959). Investment in Man: An Economist's View, *Social Service Review*, **33** (2), pp. 109–117.

——— (1961). Investment in Human Capital, *The American Economic Review*, **51** (1), pp. 1–17.

Schwarze, Johannes (1995). Simulating German Income and Social Security Tax Payments Using the GSOEP, *Cross-National Studies in Aging Program Project Paper* (19).

Scott-Clayton, Judith (2013). Information Constraints and Financial Aid Policy, in Donald E. Heller (Ed.) *Student Financing of Higher Education*, London: Routledge, pp. 75–97.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland IVC/Statistik (2015). Abiturnoten im Ländervergleich: Übersichten 2006 bis 2013, URL: http://tinyurl.com/kmkabinoten.

Sexton, Alison L. and Timothy K.M. Beatty (2014). Behavioral Responses to Daylight Savings Time, *Journal of Economic Behavior & Organization*, **107** (Part A), pp. 290–307.

Shochat, Tamar, Mairav Cohen-Zion, and Orna Tzischinsky (2014). Functional Consequences of Inadequate Sleep in Adolescents: A Systematic Review, *Sleep Medicine Reviews*, **18** (1), pp. 75–87.

Simeaner, Hans, Michael Ramm, and Christa Kolbert-Ramm (2014). Datenalmanach Studierendensurvey 1993 - 2013: Studiensituation und Studierende an Universitäten und Fachhochschulen, *Hefte zur Bildungs- und Hochschulforschung* (7).

Steele, Claude M., Steven J. Spencer, and Joshua Aronson (2002). Contending with Group Image: The Psychology of Stereotype and Social Identity Threat, *Advances in Experimental Social Psychology*, **34**, pp. 379–440.

Steiner, Viktor and Katharina Wrohlich (2012). Financial Student Aid and Enrollment in Higher Education: New Evidence from Germany, *The Scandinavian Journal of Economics*, **114** (1), pp. 124–147.

Stephens, N. M., M. G. Hamedani, and M. Destin (2014). Closing the Social-Class Achievement Gap: A Difference-Education Intervention Improves First-Generation Students' Academic Performance and All Students' College Transition, *Psychological Science*, **25** (4), pp. 943–953.

Stinebrickner, Ralph and Todd R. Stinebrickner (2003). Working During School and Academic Performance, *Journal of Labor Economics*, **21** (2), pp. 473–491.

Stock, James and Motohiro Yogo (2005). Testing for Weak Instruments in Linear IV Regression, in Donald W. K. Andrews (Ed.) *Identification and Inference for Econometric Models*, New York: Cambridge University Press, pp. 80–108.

Strauss, Robert P. (1977). Information and Participation in a Public Transfer Program, *Journal of Public Economics*, **8** (3), pp. 385–396.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization, *The Review of Economic Studies*, **23** (3), pp. 165–180.

Thaler, Richard H. and H. M. Shefrin (1981). An Economic Theory of Self-Control, *The Journal of Political Economy*, **89** (2), pp. 392–406.

Triventi, Moris (2014). Does Working During Higher Education Affect Students' Academic Progression? *Economics of Education Review*, **41**, pp. 1–13.

Valdez, Pablo, Candelaria Ramírez, and Aída García (2003). Adjustment of the Sleep-Wake Cycle to Small (1-2h) Changes in Schedule, *Biological Rhythm Research*, **34** (2), pp. 145–155.

Van Dongen, Hans, Greg Maislin, Janet M. Mullington, and David F. Dinges (2003). The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology from Chronic Sleep Restriction and Total Sleep Deprivation, *Sleep*, **26** (2), pp. 117–129.

Vandewalle, Gilles, Evelyne Balteau, Christophe Phillips, Christian Degueldre, Vincent Moreau, Virginie Sterpenich, Geneviève Albouy, Annabelle Darsaud, Martin Desseilles, Thien Thanh Dang-Vu, Philippe Peigneux, André Luxen, Derk-Jan Dijk, and Pierre Maquet (2006). Daytime Light Exposure Dynamically Enhances Brain Responses, *Current Biology*, **16** (16), pp. 1616–1621.

Vandewalle, Gilles, Pierre Maquet, and Derk-Jan Dijk (2009). Light as a Modulator of Cognitive Brain Function, *Trends in Cognitive Sciences*, **13** (10), pp. 429–438.

Van de Ven, Wynand P. M. M. and Bernard van Praag (1981). The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection, *Journal of Econometrics*, **17** (2), pp. 229–252.

Vincent-Lancrin, Stéphan (2009). The Reversal of Gender Inequalities in Higher Education: An On-going Trend, in OECD (Ed.) *Higher Education to 2030*, **1**, Demography, Washington: OECD Publishing, pp. 266–298.

Vriend, Jennifer L., Fiona D. Davidson, Penny V. Corkum, Benjamin Rusak, Christine T. Chambers, and Elizabeth N. McLaughlin (2013). Manipulating Sleep Duration Alters Emotional Functioning and Cognitive Performance in Children, *Journal of Pediatric Psychology*, **38** (10), pp. 1058–1069.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007). The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements, *Schmollers Jahrbuch*, **127** (1), pp. 139–169.

Weisbrod, Burton A. (1970). On the Stigma Effect and the Demand for Welfare Programs: A Threoretical Note, *Institute for Research on Poverty Discussion Papers* (82-70).

Whelan, Stephen (2010). The Take-Up of Means-Tested Income Support, *Empirical Economics*, **39** (3), pp. 847–875.

Whiteside, Stephen P. and Donald R. Lynam (2001). The Five Factor Model and Impulsivity: Using a Structural Model of Personality to Understand Impulsivity, *Personality and Individual Differences*, **30** (4), pp. 669–689.

Wiemers, Jürgen (2015). Endogenizing Take-up of Social Assistance in a Microsimulation Model: A Case Study for Germany, *International Journal of Microsimulation*, **8** (2), pp. 4–27.

Wilde, Joachim and Alexander Kubis (2005). Nichtinanspruchnahme von Sozialhilfe: Eine empirische Analyse des Unerwarteten, *Jahrbücher für Nationalökonomie und Statistik* (3), pp. 347–373.

Wiswall, M. and B. Zafar (2015). Determinants of College Major Choice: Identification using an Information Experiment, *The Review of Economic Studies*, **82** (2), pp. 791–824.

Wiswall, Matthew and Basit Zafar (2013). Belief Updating Among College Students: Evidence from Experimental Variation in Information, *SSRN Electronic Journal*.

Wolfson, Amy R. and Mary A. Carskadon (1998). Sleep Schedules and Daytime Functioning in Adolescents, *Child Development*, **69** (4), p. 875.

——— (2003). Understanding Adolescent's Sleep Patterns and School Performance: A Critical Appraisal, *Sleep Medicine Reviews*, **7** (6), pp. 491–506.

Wooldridge, Jeffrey M. (1995). Score Diagnostics for Linear Models Estimated by Two Stage Least Squares, in C. Radhakrishna Rao, G. S. Maddala, Phillips, P. C. B, and T. N. Srinivasan (Eds.) *Advances in Econometrics and Quantitative Economics*, Oxford and UK and Cambridge and USA: B. Blackwell, pp. 66–87.

Wößmann, Ludger (2003). Schooling Resources, Educational Institutions and Student Performance: The International Evidence, *Oxford Bulletin of Economics and Statistics*, **65** (2), pp. 117–170.

——— (2004). How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the US, *CESifo Working Paper* (1162).

Yamamoto, K. and E. Kulick (2012). TIMSS and PIRLS Achievement Scaling Methodology, in Michael O. Martin and Ina V.S Mullis (Eds.) *Methods and Procedures in TIMSS and PIRLS 2011*, Chestnut Hill and MA: TIMSS & PIRLS International Study Center, Boston College.