Microeconometric Analyses of Cognitive Achievement Production

Bernhard Enzi

ifo
BEITRÄGE
zur Wirtschaftsforschung

Microeconometric Analyses of Cognitive Achievement Production

Bernhard Enzi

Microeconometric Analyses of Cognitive Achievement Production

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München
Volkswirtschaftliche Fakultät

2016

vorgelegt von
Bernhard Enzi

Referent: Prof. Dr. Ludger Wößmann
Korreferent: Prof. Dr. Oliver Falck
Promotionsabschlussberatung: 10.5.2017

Contents

Tables

Figures

List of abbreviations

DID   Difference-in-differences

DZHW German Centre for Research on Higher Education and Science Studies

EPF   Educational production function

FE   Fixed-effects

FSE   First state examination

GPA   Grade point average

IV   Instrumental variable

NEPS   National Educational Panel Study

OLS   Ordinary least squares

RCT   Randomized controlled trial

RD   Regression discontinuity

SD   Standard deviation

SE   Standard error

SSE   Second state examination

SUR   Seemingly unrelated regression

VA   Value-added

WSBS   Within-Student Between-Subject

1 Introduction

1.1 The Role of Human Capital in Economics

There is a large and growing body of evidence supporting the now widely accepted idea that human capital is the key to success for both individuals and societies. Various individual outcomes such as labor market success (Card 1999) and health (Brunello et al. 2013; Silles, 2009) are found to be affected beneficially by human capital, as is economic growth (Hanushek, and Woessmann, 2008) and other facets of a society's overall well-being (Michalos 2008). Education is a key means of accumulating human capital and there are several reasons why government intervenes in the education sector, including credit constraints (Carneiro, and Heckman, 2002), externalities (Lochner 2011), and, with the Internet spreading knowledge throughout the world, an ever increasing non-excludability (Stiglitz 1999).

Hence, due to these market failures and its importance for growth, education policy can be effective in increasing welfare, which is why economists became interested in the determinants of education. The theoretical and methodological foundations of education research, however, have been laid rather recently in economics. Theoretical modeling in macroeconomics, for example, neoclassical growth theory, ignored human capital (Solow 1956; Swan 1956), as did early endogenous growth models (Domar 1946; Harrod 1939). Mankiw, Romer, and Weil (1990), Romer (1989), and Lucas (1988) were the first to include human capital in the respective frameworks. In microeconomics, Becker (1962), Mincer (1958), and Schultz (1961) laid the groundwork for future research in this area. Today, there are varying views in microeconomics on how to formalize education's role in economic models with each having its merits, including: human capital as increasing labor productivity (Becker 1964), human capital as multidimensional (Gardner 1974), the skill to adapt in disequilibrium situations (Nelson and Phelps 1966), the capacity to live in an

hierarchical society (Bowles, and Gintis, 1975), or as a sole signal for an individual's underlying ability (Spence 1973; Weiss 1995).

Computational and statistical capabilities have developed rapidly over the past decade. Methods mimicking experimental designs that are now standard tools were popularized in the second half of the 20th century. For panel data methods, see Anderson and Hsiao (1982); for instrumental variable (IV) approaches, see Reiersøl (1945) and Wright et al. (1928). Other methods were popularized quite recently, for example, difference-in-differences (DID) (Ashenfelter, 1978; Card, and Krueger, 1994). Regression discontinuity (RD) was invented by Thistlethwaite and Campbell (1960) but became popular in education research only recently (Imbens, and Lemieux, 2008). Card (1990) implicitly used synthetic controls, but it took some time before this approach became common among applied microeconometricians (Abadie, and Gardeazabal, 2003). Started by Rosenbaum and Rubin (1983), in the 1990s there was an extensive debate about whether matching methods are better at reducing selection bias than standard regression analysis (Dehejia, and Wahba, 1999; Smith,and Todd, 2005). At a minimum, what this debate has shown is that matching methods provide a convincing way to select the observations on which other estimation methods can later be applied. These methods are today the basis of causal inference in the economics of education.[1]

This dissertation contributes empirically in that intersection of micro- and macroeconomics by investigating determinants of education in Chapters (2), (3) and (5). Chapter (4) analyzes educational inequality that may cause economic inequality. Chapters (2) to (4) use panel data models to account for potential endogeneity while Chapter (5) employs an IV approach.

Education's empirical importance, theoretical and methodological progress, data availability, and the general expansion of economics into other fields of social science (Grossbard-Shechtman, and Clague, 2002; Hirshleifer, 1985) led economists to become increasingly interested in education itself. Early seminal works in the empirical economics of education include, for example, Mincer (1974) on the

---

[1] For a general history of econometrics see Epstein (2014); for recent developments, see Imbens and Wooldridge (2009).

returns to education, Coleman (1966) on the relationship of school inputs and student achievement, and Hanushek (1971) on teacher research.[2]

1.2   Educational Production

Bowles (1970) and Hanushek (1979) developed the educational production framework, (EPF) that formalizes economic thinking on education. The EPF is the theoretical foundation of Chapters (2) to (4) of this dissertation and it regards cognitive achievement as a commodity that can be produced:

$$y_{it} = y_t \ [F_i, S_i, X_i, \mu_i] \qquad\qquad (1\text{-}1)$$

Cognitive achievement y of individual i at time t in some cognitive dimension is a function defined by the technology $y_t \ (.)$ that translates the educational inputs into the educational outcome. The educational inputs are a vector of the entire history of all family F, school S, and all other inputs X up to time t, as well as the mental endowment μ_i .

The EPF has been investigated meticulously and certain of its properties, such as substitutability of inputs (Linden 2008), nonlinearities (Baker 2001), and dynamic decay (Sass, Semykina, and Harris 2014) have been discovered. In practice, it is mostly linear functions with additively-separable inputs that are estimated (Todd, and Wolpin, 2003), and this is the approach taken in Chapters 2 and 3 of this dissertation. Chapter 4 investigates an adjusted version of the approach to be applied to teacher grading.

One important methodological issue is how to appropriately measure "education." Early studies focused on educational attainment, for example, years of schooling (Mankiw, Romer, and Weil 1990; Psacharopoulos 1994); more recent studies focus on educational achievement measured by standardized test scores (Bishop, 1997; Heyneman, and Loxley, 1983), non-cognitive skill measures (Heckman, and Rubinstein, 2001), or other educational outcomes such as dropout rates (see Chapter 5).

---

[2] For an overview of the economics of education field, see Hanushek, Machin, and Woessmann (2016) and other volumes of the series.

One strand of literature focuses on teachers and their impact on student achievement, with findings mostly stemming from the United States. Teacher quality is found to vary substantially in that country (Rivkin, Hanushek, and Kain, 2005), while I find the variance to be substantially lower in Germany (Chapter 2). Although teachers matter, there are hardly any observable teacher traits that correlate with teacher quality (Hanushek, and Rivkin, 2010). I find, for Germany, that the high school leaving diploma and performance on one of two exams taken for teacher certification matter for teachers' effectiveness in Chapter (2), as does their experience in Chapter (3). Some work studies incentive-pay schemes for teachers, finding mostly favorable effects on student achievement (Lavy, 2009; Muralidharan, and Sundararaman, 2009).

That literature commonly employs so called value-added (VA) models in which the lagged value of the outcome variable is used as a right-hand side variable to account for the entire history of educational production up until the timing of the lag. As education takes place over the entire life-cycle, empirical analyses have to take the entire past into account. Therefore, the VA approach is a convenient way to solve the otherwise almost impossible task of providing data for individuals over their entire lifespan.

Research concentrating on the individual student and her family chiefly focuses on the intergenerational mobility of education (Björklund, and Salvanes, 2011), particularly for disadvantaged students (Betts and Roemer 2005). Although family traits are very important in the intergenerational mobility of education, they are difficult to manipulate by policy and thus other sorts of public intervention are of interest. For example, it is found that monetary and, especially, nonmonetary incentives can have a significant positive impact on students, and effects are stronger the earlier they are offered (Levitt et al. 2012).

There are several school and classroom attributes that can affect student achievement. One of these is class size, although it is far from settled whether bigger or smaller class sizes are optimum. Many studies show that smaller classes increase student achievement (Angrist, and Lavy, 1997); others find the effect to be insignificant (Hanushek, 2003). Peers have been found to significantly affect the individual student (Sacerdote 2011), but these effects are hard to identify in the absence of explicit random assignment (Angrist, 2013). Another strand of research that focuses on peer quality has moved away from measures of average peer quality to ranking students within the classroom achievement distribution and finds significant positive effects of a higher rank (Murphy, and Weinhardt, 2014).

Other studies take a broader look at the school system in general. School choice (Angrist et al. 2006), competition (Böhlmark, and Lindahl, 2015), accountability, and autonomy can be powerful tools for education policy to increase welfare (Abdulkadiroguglu et al. 2011) if put together in the right mix (Bishop, 1997; Woessmann et al. 2009). Central exit exams can provide such an environment and are found to positively affect students' academic and labor market performance (Jürges et al. 2005; Piopiunik et al. 2013). Tracked school systems are a particularly European topic and are found to have negative effects on disadvantaged students, nor do they advantage high performing students (Hanushek, and Woessmann, 2006; Kerr, et al. 2013).

Compared to secondary education, there is much less research on tertiary education. Parey and Waldinger (2011), for example, use the expansion of the ERASMUS scholarship program to investigate the causal effect of studying abroad and find that it increases the likelihood of working abroad by 15 percent. Chapter 5 contains a more extensive review of the literature on this topic and an investigation of the impact of the Bologna reform on student outcomes.

1.3   Causal Inference

Angrist and Pischke (2008) set a standard for investigating causal effects in the absence of a deliberately randomized treatment. Usually, there are several reasons for why correlations do not imply causation (e.g., omitted-variable bias, reverse causality) and why regression coefficients do not accurately measure the impact of one variable on the other (e.g., measurement error in the independent variable).[3]

Randomized control trials (RCT) produce deliberate randomized treatments and are viewed as the gold standard in causal inference. In these trials, some treatment's intensity is randomized between the participants; hence any potential confounding factors that may be of importance for the respective outcome cannot be correlated with the treatment itself and ordinary least squares (OLS) estimates consistently estimate the causal effect of the treatment. In education research, typically practical issues such as cost, ethics, self-selection into treatment, long durations, and external validity may hinder the realization or its interpretation.

---

[3] For a practical guide to standard econometric tools in the economics of education, see Schlotter, Schwerdt, and Woessmann (2011).

Other methods to estimate causal effects depend on the identifying assumption the researcher is willing to make. A natural place to start is conditional random assignment of treatment on covariates. In this case, some treatment is randomized, but only after accounting for certain factors. Conditioning on these factors either in a matching or OLS framework can solve the problem of omitted-variable bias or selection on observables, depending on the specific setting.

One main problem in observational studies in the economics of education is unobserved, time-invariant heterogeneity of observational units that is correlated with the treatment. A classic example of this is when students with a strong, but unobserved, innate ability select themselves into some treatment, for example, smaller classrooms. Ordinary least squares estimates of the effect of test score on class size would then be biased downward. One standard method for dealing with such a problem, when panel data are available, is FE estimation. De-meaning (or first differencing) the outcome and the right-hand side of the regression model removes the unobserved ability component. Hence, OLS estimates of the de-meaned model will no longer be biased by the component that was algebraically canceled out.

A special case of FE is the DID approach, which can be applied in cases where a policy affects a particular group. Assuming that this group would have, over time, developed the same as the rest of the population in the absence of the policy, this method yields a consistent estimate of the treatment effect by comparing changes in the outcome variable over time across groups.

A widely used panel data model in teacher research is the VA representation of the EPF, which can be viewed as an adjusted fixed effects (FE) model that does not de-mean or first-difference the independent variable. This model is used in Chapters (2) and (3). One crucial assumption of this model is random assignment of teachers to students conditional on a lagged measure of student achievement. However, this is a much-debated assumption, and there is some question of whether results from VA models "can be trusted" (Chetty et al. 2014; Goldhaber, and Chaplin, 2015) or not (Guarino et al. 2014; Rothstein, 2010).

The IV approach is an established technique for dealing with endogeneity of the treatment. The identifying assumption is that there exists some variable that is correlated with the variable of interest, but not with any other factor determining the outcome that cannot be controlled for. This exogenous variation, for example, may stem from governmental randomization or other unforeseeable and

unswayable conditions. An example is (high-quality) schools that are oversubscribed and base their enrollment on lotteries. The lottery's outcome can be used to instrument the endogenous treatment of attending that particular school (Angrist, Pathak, & Walters, 2013).

RD designs belong to the class of IV estimators. In an RD setting, the identifying assumption is that individuals who are (closely) above and below a cutoff (e.g., date or score) share the same unobserved attributes that may be correlated with the treatment of interest (Schwerdt, West, & Winters, 2015). IV is also a powerful way of dealing with (classical) measurement error in the independent variable that causes attenuation bias toward zero in OLS estimations. If the standard assumptions of relevance and exogeneity of the instrument are fulfilled even in the presence of measurement error, the IV estimate is consistent.

To test for statistical significance, standard errors or test statistics must be estimated appropriately. In education research there are several reasons for not assuming homoscedasticity of the error term. For example, often there is serial correlation in the error terms and no theoretical guidance on whether to assume a homo- or heteroscedastic form of errors. As the heteroscedastic estimation of standard errors is less restrictive and leads to consistent estimation even in the presence of homoscedasticity, it should be preferred. Most studies refrain from imposing any functional assumptions on the structure of the error term and use cluster-robust standard errors (preferably) on the level of randomization.

1.4   Outline of This Dissertation

This thesis contributes to the literature on determinants of student achievement. All chapters are about the educational system in Germany. Chapters 2 to 4 deal with topics in secondary education using data from the German National Educational Panel Study (NEPS), while Chapter 5 is about tertiary education using data provided by the German Centre for Research on Higher Education and Science Studies (DZHW). The impact of teachers, the role of the Bologna Reform, and the gender-specific pattern of grading students by teachers are analyzed. Conceptually, all chapters are based on the educational production framework. Methodologically, microeconometric methods to account for likely biases are employed in order to facilitate causal interpretation of findings.

Short and stylized overviews of the educational institutions relevant in the context of the studies are presented in Chapters 2.2, 3.3 and 5.3. Education policy in

Germany is set at the state level, thus complicating the discussion of a "German" education system, as there are 16 states. If not stated otherwise, I refer to all 16 states. For a detailed overview of the 16 education systems in the year of 2010, see Lohmar and Eckhardt (2010). Following is a brief summary of Chapters 2 to 5:

Evidence collected in schools in the United States reveals that principals are able to identify effective teachers once the teachers have been teaching for several years. However, the lack of observational evidence of teachers' characteristics before they enter the labor market prevents identification of effective teachers at the hiring stage. To test whether effective teachers could be identified before labor-market entry, I utilize the German setting in Chapter 2, in which prospective teachers have to take a first exam on content knowledge and a second exam where experts grade their teaching skills. I apply standard value-added models to rich German student achievement panel data. I find that performance on the pedagogical and high school leaving exam, but not on the cognitive exam, predicts teachers' effectiveness in schools. A better grade on the pre-service pedagogical exam is associated with more efficient classroom management. I also find that the distribution of teacher effectiveness is narrower in Germany than in the United States.

Chapter 3 asks the question of what makes a "good" teacher. Outside of the United States, little is known about determinants of teacher quality, and it is unlikely that determinates discovered in the U.S. context are generalizable to other parts of the globe. Using rich micro-level data from Germany, I exploit within-student, between-subject variation in mathematics and German to account for potential sorting of students and teachers. I investigate a battery of determinants from various dimensions: demographics, attitudes, education, (previous) labor market experience, and measures of intrinsic motivation. The results indicate that the previously discovered positive effect of initial years of experience holds for mathematics teachers, but not for German teachers. Furthermore, teachers in both subjects become less effective later in their careers. Other potential determinants do not have predictive power for teacher quality.

Are there systematic differences in the way teachers grade their male and female students conditional on the same performance? In Chapter 4, using rich micro-level data from NEPS and applying fixed effects estimators to account for unobserved heterogeneity, I show that girls are graded worse in mathematics, compared to equally performing boys, whereas boys are graded worse in German compared to equally performing girls. No such gender gap exists for science. The

findings are robust to several specifications and cannot be explained by observed non-cognitive skills, teacher characteristics, or in-class activities.

Finally, by combining rich data on university students with administrative data on universities' study programs, Benedikt Siegler and I exploit variation in the timing of Bachelor degree introduction across departments to investigate the impact of the Bologna Reform on students in Chapter 5. To account for endogeneity in students' enrollment decisions, we apply an instrumental-variable approach based on the distance differential between an individual's nearest universities with a Bachelor's and a traditional degree program. Overall, we find no effects of the reform on students' mobility, dropout, and internship participation, although there is indication that the reform reduced dropout for females and for high-achieving students and increased study satisfaction.

2 The Effect of Pre-Service Cognitive and Pedagogical Teacher Skills on Student Achievement
Evidence from German Entry Screening Exams[1]

2.1 Introduction

How can teachers be selected more efficiently? Teacher quality research shows that there are few determinants that help identify effective teachers in service and even fewer before service. This study increases our knowledge in this area by investigating the screening devices employed by German education policy to determine entry into the teaching profession: the grades received in two state examinations.

The study makes two contributions. I provide the first empirical evidence on the screening process for teachers in the German education system. Screening is primarily based on grade point averages from two state examinations, which are assigned by the respective federal state's education ministry during the teacher training program. As its second contribution, this is the first study to estimate the variance of teacher quality in Germany. Due to the particular German teacher training and labor market, the distribution of teacher quality is far from clear and is, moreover, dependent on the assumptions one is willing to make for how the longer and more organized screening works. While teacher quality is key for educational

efficiency, its variance is essential for educational equity and, in the long run, for economic equality.

Teacher screening is almost exclusively based on the (weighted) grade point average the applicant received in two state examinations. The first examination measures cognitive and theoretical pedagogical skills; the second examination practical pedagogical skills based, partially, on demonstration lessons graded by a head teacher. I further analyze the teachers' high school GPA (or "Abitur" grade), which can be viewed as a second-order screening device that partially determines entry into the teacher training programs at the university level. Effective selection of the teaching force is likely to be a cost-effective way to increase educational outcomes as good student-teacher matches will lead to real educational and accompanying economic gains. Furthermore, as dismissing a teacher can be costly or even judicially virtually impossible, effective screening is all the more important.

I apply a standard value-added model to very rich micro-level data from the German National Educational Panel Study (NEPS) that comprises a nationally representative sample of secondary school students to analyze the determinants and distribution of teacher quality.

I find that teacher quality is very likely to be lower in Germany than in the United States, with upper-bound estimates of 0.15 standard deviations (SD) in German and 0.13SD in mathematics. High school GPA and the grade received on the second state examination are statistically significant determinants of teacher quality. These two grades are correlated with less time needed for classroom management.

This part is organized as follows: Chapter 2.2 describes the literature on determinants of teacher quality and the institutional background regarding teacher training and the teacher labor market in Germany. The data are presented in Chapter 2.3. Chapter 2.4 presents the value-added framework and the strategy for estimating the parameters of interest. Chapter 2.4.1 discusses the estimation of the teacher quality distribution and its results. Main results for the screening variables are presented in Chapter 2.5 and their in-class time activity correlates in Chapter 2.6. Chapter 2.7 contains results for different student subgroups and Chapter 2.8 concludes.

2.2 Background

2.2.1 Literature on Teacher Effectiveness

There is a great deal of work on teacher effectiveness in the economics of education literature. For a general overview, see Hanushek, and Rivkin (2006). This literature can be divided into three subfields: estimation of teacher quality variance, investigations of determinants of quality differences, and analyses of teacher policies or interventions.

The strand of the literature that focuses on the effects of teacher education or certification is the one most closely related to this study. In the United States, this research is mainly motivated by the fact that many states provide additional compensation to teachers who hold a master's degree or have been certified by the National Board of Professional Teaching Standards. However, there appears to be little or even no impact of an advanced degree on student learning (Clotfelter, Ladd, and Vigdor 2006; D. Goldhaber 2002; Harris and Sass 2011; Rivkin, Hanushek, and Kain 2005), and the literature provides mixed evidence as to the effects of board certification. Some studies such as Cantrell et al. (2008) for Los Angeles and Ladd, Sass, and Harris (2007) for North Carolina) identify a positive correlation between effectiveness and NBPTS certification. Goldhaber and Anthony (2007), however, show that NBTPS certification does not enhance teacher effectiveness. Moreover, any observed correlation between certification and effectiveness might be driven by more effective teachers being more likely to obtain certification.

Another strand of research investigates how teachers perform in examinations themselves. Harris and Sass (2011) find no correlation between SAT performance and classroom effectiveness in Florida, while Boyd et al. (2008) find a positive correlation with mathematics SAT scores in New York City. Clotfelter et al. (2006) find a very weak correlation between licensure test performance and classroom effectiveness in North Carolina. However, several studies discover significant relationships between measures more closely attuned to the content knowledge used in teaching and student achievement (see the review in Baumert et al. (2010)). Hill, Rowan, and Ball (2005) find that teachers' mathematical knowledge is significantly related to student achievement in the first and third grades in U.S. elementary schools. Metzler and Woessmann (2012) estimate the causal effect of teacher subject knowledge on student achievement using within-teacher within-student variation, exploiting a unique Peruvian sixth-grade dataset that tested students and their teachers in two subjects. This allows circumventing omitted-

variable and selection biases by using student and teacher fixed effects and observing teachers teaching both subjects in one-classroom-per-grade schools. The findings suggest that a one standard deviation in subject-specific teacher achievement increases student achievement by about 10 percent of a standard deviation.

For Germany, the COACTIV study, which is embedded in the longitudinal component of the German PISA 2003 study, involves two assessments of students and their mathematics teachers at the end of grades 9 and 10. It provides evidence on the association between student achievement and different dimensions of teacher knowledge. Teachers were tested in mathematics-related content knowledge (conceptual and/or procedural mathematical skills) and mathematics-related pedagogical content knowledge (teachers' knowledge of tasks, student cognitions, and instruction). Results from COACTIV indicate significant relationships between student achievement and two types of teachers' knowledge: content knowledge and pedagogical content knowledge (Baumert and Kunter 2011; Baumert et al. 2010; Mareike Kunter et al. 2007).

There is also research more directly focused on the actual teaching process and the impact of in-class activities on student achievement (for reviews of this literature, see Seidel and Shavelson (2007) and Slavin, Lake, and Groff (2009). The findings of this literature underscore the importance of teaching practices, instructional skills, and classroom management for student performance. For example, Kane, Staiger, and Rockoff (2010) and Tyler et al. (2010) find that classroom management and instructional skills as measured by the Teacher Evaluation System in Cincinnati can predict student achievement. Their classroom observation measures capture teaching practices such as "the teacher establishes effective routines and procedures … and manages transitions to maximize instructional time" and "the teacher engages students in discourse and uses thought-provoking questions." Lavy (2010) finds that teaching emphasizing in-the-classroom instilment of knowledge and comprehension has a very strong and positive effect on test scores. A meta-analysis by Slavin et al. (2009) reveals significant impacts of cooperative learning programs in mathematics instruction that target teachers' instructional behaviors rather than mathematics content alone. Lou et al. (1996) argue that within-class grouping, a typical component of cooperative learning strategies, has potential to enhance student achievement. However, another meta-analysis by Dignath and Buettner (2008) reveals negative impacts of group work for primary school students. Based on data from the

National Educational Longitudinal Study, Goldhaber and Brewer (1997) find that instruction in small groups and emphasis on problem solving is associated with lower student mathematics test scores for 10th-grade students. Aslam and Kingdon (2011) analyze student achievement data for Pakistan and find that students have higher test scores when taught by teachers who spend more time on lesson planning and ask more questions in class. In his quasi-experimental evidence on class-size effects in Europe, Woessmann (2005) shows that the impact of class size on student achievement decreases with teacher quality.

2.2.2 Educational Institutions

In Germany, teaching is organized in classes, rather than by courses and, in general, all students in one classroom receive teaching from the same teacher in a given subject, thus not affecting estimation due to within-classroom tracking. Furthermore, teachers do not specialize in teaching one specific grade, but are assigned by school management to certain classes on a yearly basis. Teachers teaching a certain subject in Grade 5 are therefore quite likely to teach the same class in Grade 6.

2.2.2.1 Prospective Teachers' Transition from High School to University

Secondary education is tracked in Germany and only specific tracks give access to tertiary education. The degrees that give access to universities are all designed quite similarly and are earned during the last two years of high school and by the final examinations. Students are somewhat free to choose a set of courses and their duration and receive a final grade every semester for every class they have taken. Students additionally choose to be tested in four or five subjects by somewhat standardized final examinations. A weighted average of grades earned each semester, the grade received on one term paper, and final examinations form the high school GPA.

Conditional on having earned one degree giving access to university, enrollment in tertiary education is either open or almost exclusively based on the high school GPA.[2] After a student who wishes to become a teacher receives her high school degree, she applies to university programs that are designed to determine the school

---

[2] Some programs also give weight to motivational letters or extracurricular activities.

track, the subjects, and the state in which she will eventually teach. This is different from the U.S. system in two important ways. First, in Germany, students do not have a first year at university during which they receive general education in various fields. Second, to a large extent the German teaching program in which a student enrolls determines the future employment trajectory of a prospective teacher.

As the high school GPA includes grades received during two years of high school and those received on four or five final examinations, it is quite rich in information. It is also quite different from the U.S. SAT in that the grades included have been received for written and oral examinations, presentations, term papers, and final examinations in a broad range of subjects and thus the German high school GPA can be regarded as a measure of general education.

2.2.2.2 Teacher Training for Secondary Education

In Germany, 75 to 80 percent of teachers are graduates of a formal teacher education program (BMBF 2012). High school graduates who decide to enter a secondary education teaching program may occasionally have to fulfill certain entry requirements for a university, dependent on their high school GPA, and must choose a program that is specific to a state, a school type, and (at least) two academic subjects.

Once entered, teacher training for secondary education in Germany takes place in two steps, with the structure and content of training varying partially on the state level. Generally, the first phase takes place at a university and lasts four to six years, depending on the state. The courses include (at least) two subjects that will later be taught, pedagogics, and internships at schools. At the end of the first phase, student teachers must take exams that measure theoretical knowledge in the taught subjects and pedagogics. The outcome of these exams and the grades earned at the university level (weighted by class credits) comprise the first state examination grade.

The second stage of teacher training involves a one-and-a-half to two-year practical program of teacher seminars at teacher training schools. During this phase, every student teacher is given a teaching position. Trainee teachers are employed and teach regular classes. During this phase, trainee teachers must complete a thesis, pass several oral examinations in the subjects taught, and present three demonstration lessons that are rated by head teachers. The second state examination grade is based on the thesis grade, the oral exams, and the assessments of the demonstration lessons.

2.2.2.3 Teacher Labor Market

Entry into the profession is based on the supply of teachers and the demand for them by schools. Generally, teaching degrees specify the subjects to be taught, the type of school, and the state, and the markets are divided accordingly. For each cohort of student teacher graduates, each market clears on the basis of the (weighted) grade point average in the two state examinations.

Teachers who successfully enter the profession rarely exit before retirement. The mean leaving age for men is 60, for women 55, with medians of 62 and 60, respectively (BMBF 2012, Table 3.2.). The two most common reasons besides retirement for leaving (temporarily) are the birth of a child and long-term illness, which explains why the leaving age distribution is skewed. Most teachers are civil servants, and teacher pay is regulated at the state level, based largely on tenure and partially on assessments by principals.

2.3 Data

I use data from the German National Educational Panel Study (NEPS), which is a European education research project that began in 2010. The project draws from a representative sampling of individuals from six starting cohorts; starting cohort one (SC1) newborns, SC2 kindergarten students, SC3 fifth-grade students, SC4 ninth-grade students, SC5 university students, and SC6 adults. There are data in SC1 to SC4 from parent questionnaires. SC2 to SC4 also include data drawn from interviews with persons from other contexts in the student's life. SC2 to SC4 include data from educators' questionnaires. In the school contexts of SC3 and SC4, data are available from questionnaires answered by students, parents, and teachers of mathematics and German, as well as by principals.[3]

Figure 2-1 is a graphical representation of the data employed in this study. I use data from the first three waves of SC3 starting in 2010. Students in this cohort were sampled using a stratified sampling procedure. Schools were randomly drawn from the population of public schools to be representative by school type. From the selected schools, two classrooms (if available) were randomly asked to participate in the study. For students, participation in the study involves testing and completing a

---

[3] For a detailed overview of the NEPS, see Blossfeld, Roßbach, and von Maurice (2011); for the competence tests in particular, see NEPS (2011a, 2011b).

questionnaire. Other relevant persons (parents, homeroom teachers, principals, and teachers of mathematics and German) were also required to answer questionnaires. Participation was voluntary.

In addition to the testing information, data from student and teacher questionnaires were used for the main estimation specifications. The student questionnaires give insight into socioeconomic background and are less affected by attrition than the parental data, making them the optimal choice in the tradeoff between covariate availability and representativeness. Table 2-1 contains a selection of student background information by school track. which is later used as covariates, and also includes the outcome variables.

In contrast to administrative data from the United States, the teacher questionnaires are extensive and provide information about teachers' demographics, philosophies, educational goals, stress in the profession, colleagues, perception of the profession, participation in extracurricular activities and further training, aspects of career choice, certification, study history, subjects taught, high school GPA, and state examination grades. To make estimates comparable to other studies, I standardize the high school GPA and the grades received on the state examinations to have a mean of 0 and a SD of 1, with any of them increasing meaning a better performance. Table 2-2 provides information about basic teacher characteristics relevant for this study. For a subset of teachers, data about in-class time use is also available and will be analyzed in Section 2.5.

The outcome variables are scores on standardized tests in mathematics and language, standardized to have a mean of 0 and a SD of 1. Testing in the first wave took place in November and December 2010, near the beginning of the school year.

2.4   The Value-Added Model

The starting point of my main analyses is a slight variation of the educational production framework developed in Todd and Wolpin (2003), which describes the process of cognitive skill production as follows:

$$y_{it}^k = y_t^k[\Gamma_{TS,j}, X_i^k, \mu_i^k, \varepsilon_{it}^k] \tag{2-1}$$

Cognitive achievement y of individual i at time t in the cognitive dimension of subject k is a function defined by the technology $y_t^k(.)$ that translates the educational inputs of the vector of the entire history of all family and school (X) and teacher (Γ_{TS})

inputs until time t, as well as the subject-specific time-invariant mental endowment μ_i^k, into the educational outcome.

2.4.1 Distribution of Teacher Effectiveness in Germany

The final form of the value-added model is partly determined by available data and is similar to the empirical model of Aaronson, Barrow, and Sander (2007).The data include current test scores (Grade 7) in mathematics and language, two-year lagged test scores (Grade 5), and time-invariant family and student inputs, and allow for an analysis of the following additively-separable regression equivalent of Equation (2-1) with teacher FE τ:

$$y_{i7}^k = \lambda^k y_{i5}^k + \tau_j^k + \gamma X_i^k + \varepsilon_{i7}^k \qquad\qquad (2\text{-}2)$$

Note that for this analysis it is essential to reduce the sample to students who were taught by the same teacher in both years, as otherwise an individual teacher's contribution to educational production could not be disentangled from the other teacher's.[4] Figure 2-2 shows the distribution of the estimated teacher fixed-effects, as well as the number of teachers and the respective estimated standard deviation of teacher quality.

Note that, in contrast to common estimation results from the United States, these results are neither shrunken nor have they taken into account general classroom effects, including peer effects. Therefore, the estimates should be viewed as upper-bound estimates of the teacher quality variance in Germany. The standard shrinkage procedure as described, for example, in R. Chetty, Friedman, and Rockoff (2013a), as well as adjustments for classroom effects require multiple years of teacher observations, which are not available in my data. Generally, shrinkage leads to about a 15 percent decrease in mathematics and a 25 percent decrease in language standard deviations. (Rothstein 2010) This would result in estimates of the SD of 0.11 and 0.12, respectively, which, compared to U.S. estimates, are at the bottom of the estimate ranking (Hanushek & Rivkin, 2010) without even adjusting for classroom effects.

---

[4] If students from lower-achievement trajectories are more affected by the breaks of student-teacher matches, then, naturally, the teacher quality variance would be underestimated. However, there is no descriptive evidence that students with the same teacher in both years are different from ones with two teachers.

This finding leads to the following hypotheses. First, under signaling theory, it could be the case that German teaching degrees actually signal underlying teaching ability and the distribution of teacher applicants is cut at some threshold point, leading to a narrower distribution. Second, under human capital theory, it may be that those beginning teachers who are in the bottom part of the underlying teacher ability distribution who especially benefit from teaching programs, hence compressing the teacher quality distribution. The data at hand are not capable of determining whether signaling or human capital theory applies nor whether decreasing the teacher quality variance is more about screening teachers or educating them.

2.4.2 Pre-Service Cognitive and Pedagogical Teacher Skills

To investigate the relationship between teacher effectiveness and teacher skills, I replace the teacher FE component with the respective trait T as in:

$$y_{i7}^k = \lambda^k y_{i5}^k + \beta^k \times T_{j(5,6)}^k + \gamma X_i^k + \xi_{i7}^k \tag{2-3}$$

The test score y of student i at the beginning of *Grade 7* in the skill domain $k \in (math, language)$ is a function of its second lag y_{i5}^k, teacher j's trait T, and i's individual and family inputs. Note that the sample is limited to students who were taught by the same teacher in both Grades 5 and 6, as otherwise the teacher components cannot be distinguished.

The direct estimate of the effect of teacher characteristics on student achievement produced by Equation (2-3) is not biased by between- or within-school sorting of students based on unobservable student traits. However, to interpret the estimated vector of coefficients β causally, one would have to make the identifying assumption that unobservable teacher characteristics that directly influence student achievement are not related to the observed teacher characteristics. This is a strong assumption and one that is likely to be violated. Hence, I interpret the estimates as correlates of teacher quality, which is standard in the literature on teacher effectiveness.

One potential threat to identification may be that students on high-achievement trajectories are assigned to teachers based on observable teacher traits. However, this is unlikely, as almost all students in the sample had just switched from primary to secondary school and their parents therefore do not yet have enough knowledge about the new school environment to successfully manipulate teacher or classroom

assignment. Additionally, Fischer and Enzi (2016) provide suggestive evidence that, conditional on attending a particular school, classroom assignment is random.

Nor are the results confounded by selection into schools or by time-invariant student subject-specific ability component, as the standard VA model accounts for these aspects by including the lagged test score that captures both, as in a standard first difference specification.

2.5 Main Results

All standard errors and test statistics are based on cluster-robust standard errors at the school level—the initial stage of the stratified sampling procedure of NEPS.[5] Table 2-3 and Table 2-4 show the main regression results of math and language teachers, respectively, and their high school GPA and state examinations for the two-year gain specification. Table 2-5 shows the same results for language teachers, but for the skill domain of vocabulary in Grade 6, holding constant reading speed, comprehension, and orthography in Grade 5. If these three test scores from Grade 5 jointly cover the hypothetical Grade 5 vocabulary skill, all the exclusion restrictions of a regular VA model hold. Columns (1) through (3) are simple OLS regressions including solely one teacher grade at a time and the lagged test score from the beginning of Grade 5. Columns (4) through (6) add student covariates, and Columns (7) through (9) add Grade 5 test scores in perception speed, logic, and either math or language. Finally, Column (10) puts all teacher grades in one specification. Due to a comparably small sample, item non-response, and relatively high correlations between grades (ranging from 0.25 to 0.5) my preferred specifications are those of Columns (7) through (9). All teacher grades are standardized to have a mean of 0 and a standard deviation of 1, as are the student test scores. The estimates can be interpreted as follows: a one unit or one SD increases student gains by the estimate measured in SD.

High school GPA is an economically and statistically significant determinant of teacher effectiveness in almost all specifications and in both subjects. Table 2-3 starts out with an estimate of 0.06, decreasing to 0.04 when adding basic student covariates, which is not statistically different from the original estimate. Adding covariates without the estimate substantially changing supports my assumption of

---

[5] Clustering at the classroom level does not substantially alter the standard errors.

conditional random assignment of teachers to classrooms. Adding additional test results yields my preferred estimate of 0.035; thus a teacher having a one SD higher high school GPA is associated with an increase in student achievement of 3.5% SD.

The exact same pattern can be found in Table 2-4 for two-year gains in student language achievement, albeit with less explanatory power, as is usual in research on determinants of teacher effectiveness as, apparently, it is either harder to measure language skills or it is harder to influence them. Evidence for the first scenario can be found in the substantially lower R^2 throughout all specifications compared to the math results. Less persistence in language skills is another, but unlikely, explanation.

Somewhat surprisingly, the pattern of the high school GPA being a strong determinant reappears for one-year gains. It may be that as I do not have to restrict the sample to students who were taught by the same teachers in both years, I gain some statistical power through more observations. This finding is probably not due to less measurement error or higher persistence in this skill domain because the coefficient of determination is not substantially different from the previous one. Even with conditioning on a teacher's state exam grades, the high school GPA preserves its strong predictive power for student achievement.

In this simple specification, it is solely the high school GPA of both subjects' teachers and the second state examination of math teachers that play a role in determining student achievement. Adding basic student covariates does not alter the results substantially, but reduces standard errors. Further refining the estimation procedure by adding test scores in logic, perceptual speed, ICT, and math or language further decreases the standard errors but does not alter the point estimates, consistent with random assignment of teachers conditional on the lagged test score. As found in other studies, the impact of determinants is less significant for language teachers than for math teachers. A one-point increase in a math teacher's Abitur grade increases student learning by 6 percent of a standard deviation, while a one-point increase in the second state examination increases it by 4 percent of a standard deviation.

The first state exam's (FSE) grade is almost always an economically and statistically insignificant determinant of student achievement, in line with the mixed findings for SAT scores, which are also comprised of one-shot, large-scale exams, but in contrast to findings about teacher subject knowledge by Baumert et al. (2010) and Metzler and Woessmann (2012). The first state exam, however, is not solely based on

a teacher's knowledge about one subject, but on knowledge of two or three and also includes theoretical pedagogical knowledge, hence potentially adding measurement error that leads to attenuation bias toward zero. It appears that the weighted average of the FSE is not a good determinant by itself for student achievement in either subject and regardless of the gain specification.

The results for the second state exam (SSE) are insignificantly different from zero, but lean toward a positive relationship. As the teachers were asked last about their SSE, it is the most affected by item non-response, for various possible reasons. It may be that teachers only knew their overall state exam grade and put this down for the FSE, it may be that they were discouraged by the length of the questionnaire, or it may be that the questions about grades were perceived as too intimate an inquiry, a subject about which the respondents felt uncomfortable or defensive. Hence, the results are mixed, but point toward a positive relationship.

2.6   Pre-Service Exams and In-Class Time Use

The results found for good grades in Abitur and on the SSE may be because these teachers act substantially different in the classroom than their lower-scoring counterparts. Hence, correlations of pre-service exam results and in-class time use may shed some light on the underlying forces at work. Simple regression results are presented in Table 2-6 for a subset of math and language teachers who were, in addition to the basic questionnaire, asked about the share of time they commit to certain activities within the classroom. Column (1) shows the result for the share of time spent discussing homework, Column (2) for teacher presentations, and Columns (3) and (4) for tasks with and without assistance, respectively. Column (5) represents time spent on repetitive drills, Column (6) on taking tests, Column (7) on classroom management, and Column (8) on other activities. There are no significant results with the exception of better teachers, in terms of Abitur and SSE grades, needing less time for classroom management. Hence, more effective teachers seem to be so because they need less time for classroom management and hence have more time for other activities, which are, in turn, classroom specific and not of the "one size fits all" variety and hence in contrast to the findings of Schwerdt, and Wuppermann (2011) that find traditional teaching to be beneficial for students.

2.7   Heterogeneous Effects and Nonlinearities

Some subgroups of students have been shown to especially benefit from certain teacher traits (Dee, 2005), and girls have been shown to benefit from a single-sex

classroom in mathematics. Hence, it may be that the main results are driven by some teachers adjusting better to girls' needs in mathematics class. Table 2-7 and Table 2-8 show results for boys versus girls and pupils with a migration background versus natives, respectively.

Table 2-7 makes it clear that girls gain less in mathematics over time than do boys (see line labeled "Female"). If anything, it seems that girls benefit more from teachers with higher grades in high school and on the FSE, whereas boys benefit more from teachers with higher teaching capabilities as measured by the SSE.

Table 2-8 reveals that students with a migration background tend to gain less over time. Adding interaction terms of migration status and teachers' grades yields no statistically significant findings for migrants, although there is a tendency toward a negative correlation. However, inclusion of the migration status interaction reveals positive and significant findings for the native population for Abitur, FSE, and SSE, with SSE having the most consistent such effect. Natives are more affected by the SSE than are migrants: A one SD increase in a teacher's grade in SSE leads to 3.2% higher annual gains in math test scores for them.

I take a deeper look at the discovered effects of the state examinations and teachers' high school GPA by dividing these variables into quartiles and searching for potential nonlinearities, which are likely according to Jacob and Lefgren (2013), who found that principals can more easily tell who are the best and worst teachers, but are not doing so well at telling mediocre teachers apart from each other. Table 2-9 shows that the results in mathematics stem from the top teachers: even the FSE turns significant with a top teacher, leading to a 7.7% increase in student achievement. Hence all three grades are good at distinguishing the best from the rest when it comes to mathematics teachers. No such pattern exists for language teachers.

2.8 Conclusion

This study investigates the distribution and determinants of teacher quality in Germany, in particular pre-service skill measures for teachers in the form of two state examinations and the high school leaving diploma using rich micro-level panel data from a nationally representative sample of secondary school students starting in Grade 5. To account for potentially endogenous sorting of students into schools and classrooms, I employ a value-added model that controls for a student's previous performance within the same skill domain.

Overall, I find that teacher quality variance is lower in Germany than in the United States, consistent with either a tighter screening process or decreasing gains from these programs over the student teacher quality distribution. Students' mathematics gains are more likely to have predictors than are their language gains, as is common in the test score literature.

I find no impact from the FSE, a measure of theoretical knowledge in the teacher's subjects and pedagogics, on student gains, with the exception being the best quartile of teachers on math test scores. The second state examination is somewhat predictive of student achievement in mathematics, also stemming from the highest quartile of teachers. The high school leaving diploma is the most powerful and consistent predictor of teacher effectiveness.

The German teacher-hiring process provides a setting for a regression discontinuity approach that can be used for further research, as teachers in a given subject, state, and school track are hired up to a certain threshold each year. Longer and larger panel datasets would yield more credible results for the teacher quality variance, could be used to disentangle classroom from teacher effects, and would allow for standard shrinkage procedures.

One potential problem of teacher research in Germany is that teacher-classroom matches, which are endogenous to student achievements, are not—as in the U.S. system—limited to one year, but hypothetically can continue until the end of high school.

.

Figure 2-1: Timing of testing and available teacher data
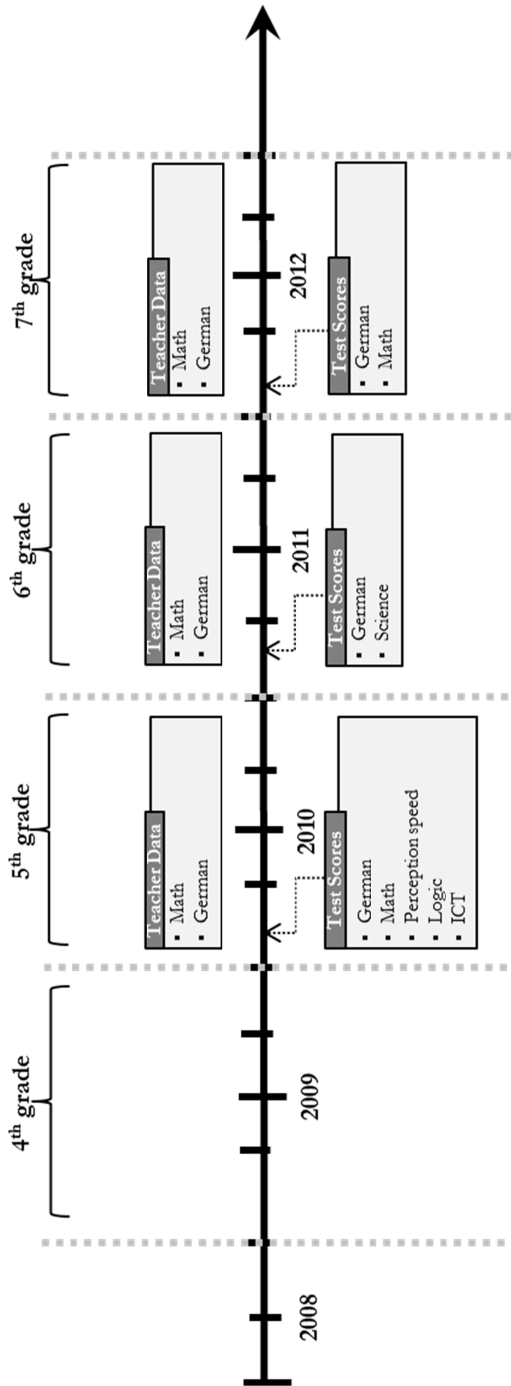
Figure 2-2: Teachers' distribution of VA by subject taught
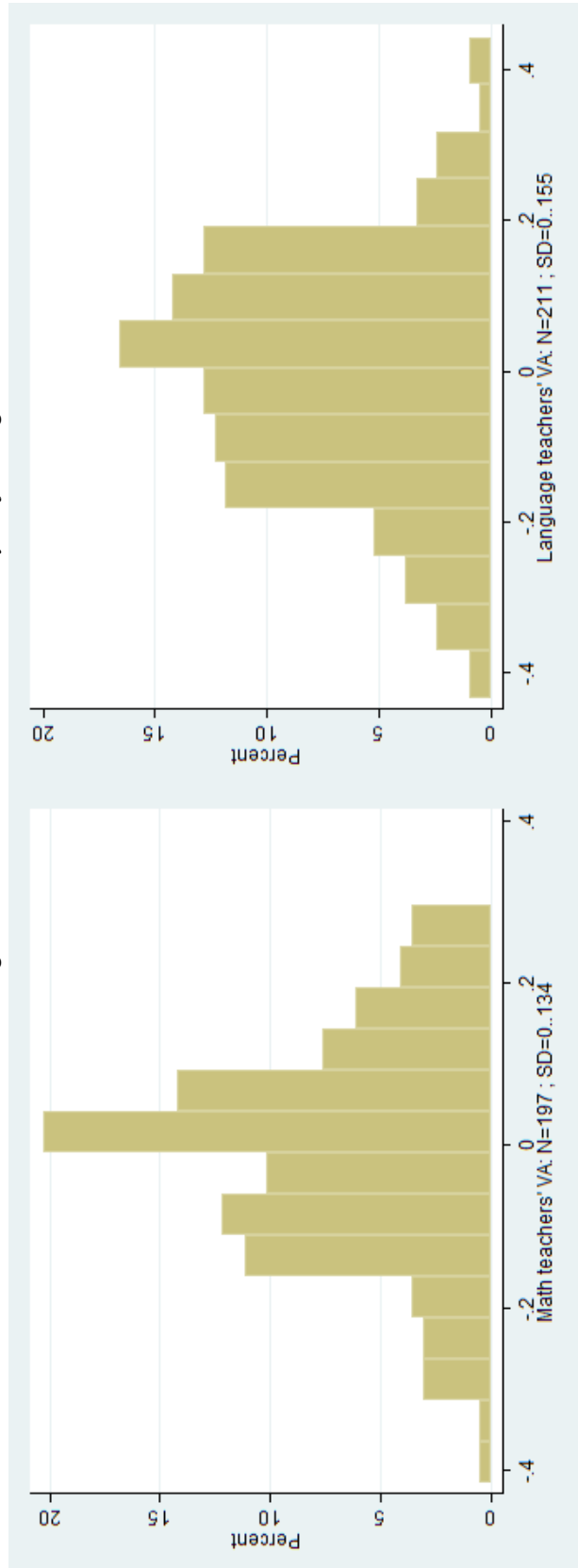
Table 2-1: Descriptive statistics for students by type of school

	Full	ES	CO	LO	ML	MI	HI
Female	0.49	0.53	0.57	0.46	0.45	0.49	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age	10.76	10.57	10.85	11.02	10.97	10.85	10.62
	(0.51)	(0.51)	(0.53)	(0.62)	(0.56)	(0.45)	(0.42)
Born abroad	0.04	0.03	0.03	0.07	0.06	0.04	0.03
	(0.20)	(0.17)	(0.17)	(0.25)	(0.23)	(0.19)	(0.18)
Household size	4.65	4.38	4.79	5.18	4.47	4.67	4.53
	(1.73)	(1.66)	(2.06)	(2.24)	(1.82)	(1.65)	(1.47)
Mother born abroad	0.19	0.14	0.36	0.25	0.12	0.18	0.17
	(0.39)	(0.35)	(0.48)	(0.43)	(0.32)	(0.38)	(0.38)
Father born abroad	0.19	0.15	0.36	0.29	0.15	0.19	0.16
	(0.39)	(0.36)	(0.48)	(0.45)	(0.36)	(0.39)	(0.37)
PC availability	1.70	1.65	1.65	1.74	1.73	1.68	1.69
	(0.52)	(0.58)	(0.58)	(0.59)	(0.60)	(0.49)	(0.48)
Retention	0.08	0.08	0.14	0.26	0.13	0.08	0.02
	(0.28)	(0.27)	(0.35)	(0.44)	(0.33)	(0.27)	(0.13)
Number of books	3.77	3.69	3.54	2.96	3.34	3.53	4.28
	(1.38)	(1.32)	(1.52)	(1.29)	(1.29)	(1.29)	(1.28)
Perceptual speed score	43.45	38.75	46.43	41.00	41.39	44.84	44.40
	(13.05)	(11.74)	(11.57)	(12.70)	(13.00)	(13.22)	(13.10)
Reasoning score	6.94	6.38	6.23	4.90	6.81	6.66	7.91
	(2.57)	(2.53)	(2.55)	(2.43)	(2.56)	(2.39)	(2.23)
German: grade point	2.32	2.17	2.69	3.22	2.61	2.57	1.85
	(0.88)	(0.83)	(0.85)	(0.85)	(0.92)	(0.71)	(0.64)
Math: grade point	2.27	2.31	2.60	3.19	2.49	2.46	1.80
	(0.94)	(0.87)	(0.86)	(1.05)	(1.00)	(0.77)	(0.68)
German: test score	0.00	-0.38	-0.35	-0.46	-0.07	-0.07	0.31
	(1.00)	(0.70)	(0.71)	(0.94)	(1.17)	(1.01)	(0.94)
Math: test score	0.00	-0.18	-0.68	-0.97	-0.32	-0.18	0.59
	(1.00)	(0.91)	(0.96)	(0.76)	(0.93)	(0.74)	(0.84)
Observations	2234	164	105	314	199	504	948

Means of variables with standard deviations in parentheses for the full sample in the first column. Each following column represents a particular school type. ES indicates elementary schools, LO the lowest secondary school track, ML secondary school track that mix middle and lowest tracks, MI the middle secondary school track, CO comprehensive schools, and HI the highest secondary school track.

Table 2-2: Basic descriptive statistics of teachers by subject taught

	Math	German	Full
		N=730	
Female	0.69	0.79	0.74
	(0.46)	(0.40)	(0.44)
Born before 1950	0.05	0.04	0.05
	(0.22)	(0.21)	(0.21)
Born in 1950s	0.33	0.31	0.32
	(0.47)	(0.46)	(0.47)
Born in 1960s	0.23	0.22	0.22
	(0.42)	(0.41)	(0.42)
Born in 1970s	0.24	0.28	0.26
	(0.43)	(0.45)	(0.44)
Born after 1979	0.16	0.15	0.15
	(0.37)	(0.35)	(0.36)
Mig. background	0.06	0.05	0.05
	(0.24)	(0.21)	(0.22)
East German	0.20	0.19	0.19
	(0.40)	(0.39)	(0.40)
Imm. start studies	0.66	0.64	0.65
	(0.47)	(0.48)	(0.48)
Prev. teaching exp.	0.74	0.69	0.72
	(0.44)	(0.46)	(0.45)
Other field of stud.	0.20	0.20	0.20
	(0.40)	(0.40)	(0.40)
Grade: Abitur	2.75	2.64	2.70
	(0.63)	(0.58)	(0.61)
Grade: 1st state exam	2.87	2.96	2.91
	(0.53)	(0.55)	(0.54)
Grade: 2nd state exam	2.94	2.99	2.96
	(0.59)	(0.58)	(0.59)

Unit of observation: teachers. Means of variables with standard deviations in parentheses. Minima and maxima for the full sample in the last column.

Table 2-3: Main regression results of math teachers' pre-service cognitive and pedagogical skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade: Abitur	0.063*** (0.023)			0.044** (0.021)			0.035* (0.019)			0.026 (0.023)
Grade: 1st state exam		0.036 (0.022)			0.022 (0.020)			0.018 (0.017)		-0.011 (0.026)
Grade: 2nd state exam			0.036** (0.018)			0.023 (0.015)			0.025* (0.013)	0.018 (0.024)
Student TS math	0.733*** (0.020)	0.739*** (0.021)	0.730*** (0.022)	0.573*** (0.021)	0.564*** (0.021)	0.558*** (0.022)	0.398*** (0.024)	0.387*** (0.024)	0.384*** (0.026)	0.378*** (0.029)
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Observations	2329	2383	2128	2329	2383	2128	2328	2381	2127	1742
R^2	0.535	0.536	0.520	0.582	0.586	0.572	0.622	0.625	0.610	0.609

Dependent variable: math test score in Grade 7. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-4: Main regression results of language teachers' pre-service cognitive and pedagogical skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade: Abitur	0.049* (0.026)			0.013 (0.025)			-0.016 (0.023)			0.004 (0.034)
Grade: 1st state exam		0.001 (0.030)			-0.016 (0.026)			-0.024 (0.023)		0.005 (0.036)
Grade: 2nd state exam			-0.012 (0.028)			-0.035 (0.024)			-0.033 (0.020)	-0.051* (0.026)
Student TS r. comp	0.630*** (0.022)	0.637*** (0.021)	0.636*** (0.024)	0.522*** (0.021)	0.518*** (0.021)	0.508*** (0.024)	0.319*** (0.023)	0.309*** (0.024)	0.316*** (0.024)	0.331*** (0.025)
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Observations	1939	2083	1701	1939	2083	1701	1937	2079	1698	1428
R^2	0.385	0.390	0.384	0.426	0.434	0.432	0.497	0.509	0.500	0.495

Dependent variable: German test score in Grade 7. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-5: Main regression results of language teachers' pre-service cognitive and pedagogical skills for one-year gains

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade: Abitur	0.133***			0.089***			0.069**			0.086***
	(0.035)			(0.030)			(0.029)			(0.033)
Grade: 1st state exam		0.056**			0.035			0.026		-0.006
		(0.027)			(0.021)			(0.020)		(0.027)
Grade: 2nd state exam			0.066***			0.028			0.026	0.006
			(0.023)			(0.022)			(0.021)	(0.026)
Student TS r. speed	0.094***	0.107***	0.095***	0.078***	0.087***	0.079***	0.058***	0.073***	0.061***	0.071***
	(0.022)	(0.022)	(0.022)	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)	(0.019)	(0.019)
Student TS orthogra.	0.013	0.006	-0.003	-0.004	-0.013	-0.031	-0.061***	-0.072***	-0.080***	-0.065**
	(0.021)	(0.023)	(0.022)	(0.021)	(0.022)	(0.023)	(0.022)	(0.023)	(0.022)	(0.025)
Student TS r. comp	0.499***	0.501***	0.511***	0.411***	0.403***	0.408***	0.280***	0.266***	0.279***	0.288***
	(0.026)	(0.025)	(0.027)	(0.022)	(0.020)	(0.023)	(0.022)	(0.021)	(0.024)	(0.027)
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Observations	2285	2426	1963	2285	2426	1963	2282	2421	1959	1619
R^2	0.329	0.309	0.317	0.454	0.451	0.450	0.492	0.491	0.488	0.494

Dependent variable: German test score in Grade 6. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-6: OLS regression results for the relationship of in-class time use and teachers' skills.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math teacher sample								
Grade: Abitur	-0.372	0.175	0.400	-1.250	1.152	0.843*	-1.327***	-0.331
	(0.608)	(0.769)	(1.014)	(1.104)	(0.803)	(0.434)	(0.417)	(0.627)
Grade: 1st state exam	0.337	-1.719**	0.934	0.900	0.182	-0.194	0.333	-0.483
	(0.560)	(0.848)	(0.952)	(1.064)	(0.836)	(0.420)	(0.428)	(0.551)
Grade: 2nd state exam	0.602	-0.200	-0.696	-0.595	0.078	0.154	0.005	0.448
	(0.558)	(0.704)	(0.986)	(1.180)	(0.814)	(0.429)	(0.413)	(0.565)
German teacher sample								
Grade: Abitur	-0.471	0.766	0.042	-0.383	-0.316	0.382	-0.647	0.564
	(0.564)	(0.579)	(0.715)	(1.026)	(0.695)	(0.436)	(0.398)	(0.434)
Grade: 1st state exam	-0.011	-0.351	0.370	0.921	-0.201	0.528	-0.451	-0.575
	(0.645)	(0.654)	(0.655)	(1.063)	(0.654)	(0.410)	(0.407)	(0.465)
Grade: 2nd state exam	-0.205	-0.773	-0.281	1.221	0.320	0.584	-0.735**	-0.766*
	(0.617)	(0.637)	(0.582)	(0.889)	(0.645)	(0.369)	(0.369)	(0.447)

Dependent variables: one dimension of in-class time use as a share of overall time available for teaching. Standard errors in parentheses. Each cell represents a distinct regression result. Column (1) covers time spent discussing homework, Column (2) teacher presentations, Columns (3) and (4) tasks with and without assistance, respectively. Column (5) represents repetitive drills, Column (6) taking tests, Column (7) classroom management, and Column (8) other activities.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-7: Heterogeneous effect of math teachers' pre-service cognitive and pedagogical skills by student gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	-0.121***	-0.122***	-0.140***	-0.143***	-0.150***	-0.173***	-0.177***	-0.176***	-0.206***	-0.176***
	(0.029)	(0.028)	(0.029)	(0.029)	(0.026)	(0.027)	(0.029)	(0.026)	(0.027)	(0.032)
Grade: Abitur	0.053**			0.031			0.022			0.031
	(0.026)			(0.022)			(0.019)			(0.025)
Grade: 1st state exam		0.030			0.009			0.002		-0.048
		(0.031)			(0.027)			(0.024)		(0.032)
Grade: 2nd state exam			0.051**			0.029			0.030	0.047
			(0.024)			(0.021)			(0.019)	(0.030)
F x Abitur	0.024			0.026			0.027			-0.007
	(0.027)			(0.025)			(0.026)			(0.031)
F x 1st StE		0.019			0.026			0.031		0.076**
		(0.032)			(0.030)			(0.028)		(0.038)
F x 2nd StE			-0.025			-0.012			-0.010	-0.059
			(0.028)			(0.027)			(0.026)	(0.038)
L. TS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	No	No	No	No	No	No	Yes	Yes	Yes	Yes
N	2329	2383	2128	2329	2383	2128	2328	2381	2127	1742
R2	0.54	0.54	0.52	0.58	0.59	0.57	0.62	0.63	0.61	0.61
Full effect	0.077	0.048	0.026	0.057	0.035	0.016	0.048	0.033	0.020	
p	0.008	0.037	0.208	0.042	0.124	0.409	0.062	0.095	0.260	

Dependent variable: Math test score in Grade 7. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-8: Heterogeneous effect of math teachers' pre-service cognitive and pedagogical skills by students' migration background

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Migrat. backg.	-0.117***	-0.108***	-0.083*	-0.044	-0.030	-0.014	-0.037	-0.029	-0.017	-0.046
	(0.045)	(0.041)	(0.043)	(0.036)	(0.034)	(0.036)	(0.033)	(0.031)	(0.033)	(0.037)
Grade: Abitur	0.059**			0.036			0.029			0.007
	(0.026)			(0.023)			(0.020)			(0.025)
Grade: 1st state exam		0.042**			0.028			0.024		-0.006
		(0.021)			(0.017)			(0.015)		(0.023)
Grade: 2nd state exam			0.040**			0.029*			0.032**	0.028
			(0.019)			(0.016)			(0.015)	(0.025)
MB x Abitur	0.023			0.023			0.014			0.074*
	(0.040)			(0.035)			(0.031)			(0.037)
MB x 1st StE		-0.023			-0.031			-0.032		-0.029
		(0.044)			(0.039)			(0.033)		(0.050)
MB x 2nd StE			-0.018			-0.034			-0.034	-0.047
			(0.039)			(0.031)			(0.029)	(0.044)
L. TS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	No	No	No	No	No	No	Yes	Yes	Yes	Yes
N	2313	2368	2118	2313	2368	2118	2312	2366	2117	1732
R2	0.54	0.54	0.52	0.58	0.59	0.57	0.62	0.62	0.61	0.61
Full effect	0.082	0.019	0.022	0.059	-0.003	-0.005	0.044	-0.009	-0.003	
p	0.036	0.695	0.538	0.084	0.943	0.868	0.133	0.804	0.915	

Dependent variable: Math test score in Grade 7. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2-9: Non-linearities in teachers' pre-service cognitive and pedagogical skills

	Math			Language		
	(1)	(2)	(3)	(1)	(2)	(3)
Abitur - 2nd Qt	0.081			0.025		
	(0.060)			(0.061)		
Abitur - 3rd Qt	0.071			0.072		
	(0.050)			(0.055)		
Abitur - 4th Qt	0.121**			-0.013		
	(0.048)			(0.056)		
1st StE - 2nd Qt		0.023			-0.001	
		(0.043)			(0.052)	
1st StE - 3rd Qt		0.010			-0.089	
		(0.054)			(0.060)	
1st StE - 4th Qt		0.077*			-0.015	
		(0.041)			(0.057)	
2nd StE - 2nd Qt			0.002			0.039
			(0.046)			(0.069)
2nd StE - 3rd Qt			0.007			-0.063
			(0.062)			(0.078)
2nd StE - 4th Qt			0.073*			-0.085
			(0.044)			(0.062)
Student TS	0.425***	0.414***	0.410***	0.323***	0.316***	0.316***
	(0.024)	(0.024)	(0.025)	(0.024)	(0.024)	(0.023)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Other tests	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2329	2383	2128	1939	2083	1701
R^2	0.618	0.622	0.607	0.500	0.511	0.505

Dependent variable: math test score in Grade 7. Standard errors in parentheses. Controls include a student's gender, year, month, and place of birth, parents' place of birth, PC availability, grades in math and language, peers' attitude toward education, school and classroom share of migrants, household size, grade retention status, number of books at home, religious denomination, and SDQ scores. Other tests include Grade 5 tests in perception speed, logic, ICT, and math or German test scores, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3 Determinants of Teacher Effectiveness

Within-Student Between-Subject Evidence from Germany[1]

3.1 Introduction

What makes a good[2] teacher? A large and still growing body of literature has investigated the distribution and potential determinants of teacher quality.[3] Teacher quality, as evaluated by raising standardized test scores,[4] has been found to vary substantially, although most investigated determinants either lack economic or statistical significance. Furthermore, most studies stem from the educational context of the United States, which has a unique institutional setting in terms of schooling in general and of teacher training and teacher labor markets in particular.

---

[2] In the economics of education literature, teacher effectiveness, productivity or quality (I use these terms interchangeably) is mostly formalized by the teacher's average ability to raise scores in standardized tests other things being equal. For an overview of the so called "value-added" literature see Koedel, Mihaly, and Rockoff (2015).

[3] For overviews in the economics of education literature see Hanushek and Rivkin (2006) and Hanushek and Rivkin (2010), for an overview of the pedagogical literature see the internationally known meta-meta-analysis of Hattie (2013).

[4] Using standardized test scores to evaluate teacher quality is appropriate despite their ordinal nature because these test scores are themselves important determinants of school attainment, earnings, and economic outcomes (Chetty, Friedman, and Rockoff, 2013; Hanushek, 2011; Hanushek, and Woessmann, 2008).

I analyze a battery of potential determinants in the domains of demographics, aspects of career choice, participation in further training, previous labor market experience, teaching philosophy, interaction with colleagues, goals for students' education, and causes of stress. As teachers in Germany rarely leave before retirement, the setting provides an interesting case study of the effects of long-term on the job training: I find that the previously found positive effect of initial years of experience exists for mathematics teachers, while it doesn't for language teachers. Both subjects, however, reveal diminishing rates of experience over time turning the mathematics effect insignificant and the language effect negative. There is further evidence for the positive effect of traditional teaching methods. Furthermore, female and teachers with migration background tend to perform worse than their male and native counterparts respectively. Secondary findings indicate that selection into classrooms by students and teachers, within a given school type, induces solely small amounts of bias, indicating a rather egalitarian distribution of teachers on observables.

The data analyzed is taken from the first wave of the German National Educational Panel Study (NEPS) starting cohorts three (about 5,000 5th graders most either 11 or 12 years of age) and four (about 15,000 9th graders most either 15 or 16 years of age). The NEPS data contains a national representative sample of schools. Within each school, two classes were randomly selected to participate in the testing. The test data focuses on language and mathematics skills and can be linked to specific classroom teachers and their responses in the survey.

The theoretical foundation of my analysis is based the notion that education is a 'classical' production process with inputs, as outlined in the educational production framework of Todd and Wolpin (2003). In the final model specification, the outcomes – test scores in math and language – are a linear, additively separable function of family, school, individual and teacher inputs. To estimate the effects of the determinants of teacher quality, I rely on the work of Dee (2005, 2007); Metzler and Woessmann (2012); and Schwerdt and Wuppermann (2011), using a generalized fixed-effects technique described by Chamberlain (1982) exploiting within-student between-subject variation to control for unobserved student heterogeneity and potential sorting to schools, classrooms and teachers.

This study is organized as follows: Chapter 3.2 summarizes the literature on determinants of teacher quality that is not included in Chapter 2.2.1. Chapter 3.3 presents the institutional setting of the German education system that is relevant for this study. In Chapter 3.4, the data is presented. Chapter 3.5 presents the educational

production framework and the estimation strategy to estimate the parameters of interest. Results are presented in Chapter 3.6 and Chapter 3.7 concludes.

3.2   Literature

The most commonly investigated teacher characteristics include basic traits (e.g., gender, experience, and education), test results (e.g., SAT performance or pedagogical and subject content knowledge), and classroom behaviors (e.g., teaching practices, classroom management).

Basic teacher demographics such as gender, race, and ethnicity are usually not viewed as main determinants of average student performance, but subgroups of students might benefit from having a demographically similar teacher. Dee (2005, 2007) investigates such student-teacher interactions. His analyses indeed suggest that racial, ethnic, and gender dynamics matter. In particular, gender interactions between teachers and students have significant effects on student test scores, teacher perceptions, and student engagement with academic subjects.

Teacher experience is one of the most investigated traits studied. (Aaronson et al., 2007; Chingos, and Peterson, 2011; Clotfelter et al., 2006; Rivkin et al., 2005; Rockoff, 2004). However, the results of these studies are mixed and indicate that teacher experience has a limited ability to explain the overall variation of teacher effectiveness. It is generally found that experience has some positive returns in the earlier years of a teacher's career, but no substantial effects are found for mid- or late-career teachers. Rockoff (2004), for example, finds marginal returns in the initial years of experience but no evidence for additional returns after five years. Rivkin et al. (2005) conclude that improvements do not seem to prevail after the first three years of teaching. Furthermore, Clotfelter et al. (2006) find that about half of the return of having an experienced teacher, which is estimated at roughly one tenth of a standard deviation in test scores, derives from the first one or two years of teacher experience.

Moreover, teachers' diagnostic competence, attitudes, and pedagogical beliefs, as well as motivation, are found to affect student achievement (Anders et al., 2010; Kunter, 2011; Voss et al., 2011).

3.3   Secondary schooling in Grade 5 and 9

After four years of elementary schooling (in Berlin and Brandenburg after six years) students enter one of the secondary school tracks. The track is determined, depending

on the state, by either a binding teacher recommendation or the parental choice. Grade 5 therefore refers to the first year of secondary education, grade 9 to the fifth year of secondary education. Depending on the state these are the lower-secondary track (Hauptschule), the middle-secondary track (Realschule), a mix of the lower and middle track, the upper-secondary track (Gymnasium) and comprehensive schools, that either do not formally track, or practice some form of within-school without within-classroom tracking. Every state has schools that exclusively offer the upper secondary school track degree and all states provide all degrees of the three tracks, the difference being the way students are divided into different school institutions. For further information about institutions in the context of secondary education see Chapter 2.2.2.

3.4 Data

For this study, I use data gathered from the first wave of SC3 and the first wave of SC4 – both from 2010. Students in these cohorts were sampled using a stratified sampling procedure. Schools were randomly drawn from the range of public schools to be representative of school type. From the selected schools, two classrooms (if available) were randomly asked to participate in the study. For students, participation in the study involves testing and completing a questionnaire. Questionnaires for relevant context persons (parents, homeroom teachers, principals and teachers in mathematics and German) were also required. For each student and his or her context persons, participation is voluntary.

In addition to the testing data, data from student and teacher questionnaires was also used for the main estimation specifications. The student questionnaires give insight into socio-economic background and are less affected by attrition than the parental data, thus being the optimal choice in the tradeoff between covariate availability and representativeness. The teacher questionnaires are also extensive and provide information about teachers' demographics, philosophies, educational goals, stress in the profession, colleagues, perception of the profession, participation in extracurricular activities and further training, aspects of career choice, certification, educational history, subjects taught, high school GPA and state examination grades.

The outcome variables are scores in standardized tests in mathematics and language, standardized to have a mean of zero and a standard deviation of one within each cohort. Testing took place in November and December of 2010, the exact date being unknown, therefore the estimates have to be considered as middle term effects

amounting to either a third or half of a school year. I restrict the data to students who are not in any kind of special education, are not taught by more than one teacher in a given subject, and can be merged with both mathematics and German teacher identifiers.[5]

Table 3-1 and Table 2-1 report descriptive statistics for students by starting cohort and type of school. The set of variables includes the outcomes (test scores in mathematics and language skills), as well as the main covariates that are used in the main specifications. The first column of Table 2-1 shows the means and standard deviations of each covariate. In total there are 2,234 observations from the younger cohort with student background information that can be linked to both teacher identifiers. The second column shows the means for students in elementary schools. This data comes from Berlin or Brandenburg and are not yet tracked. The third column shows comprehensive schools in Saxony, Schleswig-Holstein, Thuringia and Berlin in 2010. However, comprehensive schools are not the only type of school available for students of that age. Columns four to seven show, in ascending order, the tracked secondary school types. More girls attend the more selective schools. They are younger on average, less likely to be born abroad or be second generation migrants, and are much less likely to have ever been held back. They also come from higher SES backgrounds and have better grades and test scores. The columns in Table 3-1 follow the same pattern, however there are no elementary schools after nine years of schooling in any German state. The same ascending pattern of students' SES and performance is visible.

The explanatory variables of interest are teacher traits. Table 3-2 reports some of the main investigated variables. For this approach, variation across subjects is necessary. The last column in Table 3-2 also reports the difference across subjects over the respective trait and its standard deviation, showing enough variation for the respective variables across subjects to apply a within-student between-subject approach.

---

[5] As participation for each of these individuals was voluntary, about 300 (700) out of 5,500 (15,200) students from 5th (9th) grade did not participate in the testing and further 1,800 (7,500) cannot be merged with information on <u>both</u> of their mathematics and language teachers. Thus sample selection might be an issue as about one third and one half of the respective cohort's students cannot be merged with both of their teachers. I can show that on average, the students in starting cohort three are slightly worse in standardized testing, while the opposite holds for starting cohort four.

3.5 Educational production

The starting point of my analysis builds a slight variation of the educational production framework as developed in Todd and Wolpin (2003), that describes the process of cognitive achievement production as follows:

$$y_{it}^k = y_t^k[\Gamma_{TS,j}, X_i^k, \mu_i^k, \varepsilon_{it}^k] \tag{3-1}$$

Cognitive achievement of individual i at time t in the cognitive dimension of subject k is a function defined by the technology $y_t^k(.)$ that translates the educational inputs of the vector of the entire history of all family and school (X) and teacher (Γ_{TS}) inputs until time t, as well as the subject specific time-invariant mental endowment μ_i^k into the educational outcome.

The data includes current test scores in mathematics and language as well as current family and student inputs, and allows for an analysis of the following additively-separable regression equivalent of equation (1) as follows:

$$y_{it}^k = \beta^k \times T_{it}^k + \gamma_X^k X_{it} + \mu_i + \varepsilon_{it}^k \tag{3-2}$$

Current achievement is now a function of observed teacher T, school and family characteristics X and the subject-invariant ability component μ. Therefore, the error term ε^k comprises the entire history of past inputs, the subject varying ability components and the unobserved current family, teacher and school factors.

Following Chamberlain (1982), I model the correlations of the unobserved individual effect μ_i with observed inputs in the following way:

$$\mu_i = \eta_1 T_{it}^k + \eta_2 T_{it}^{-k} + \theta X_{it} + u_i \tag{3-3}$$

Substituting equation (3) in (2) yields the following correlated random effects models:[6]

$$y_{it}^{mat} = (\beta^{mat} + \eta_1) \times T_{it}^{mat} + \eta_2 \times T_{it}^{ger} + (\gamma^{mat} + \theta)X_{it} + u_i + \varepsilon_{it}^{mat} \tag{3-4}$$

$$y_{it}^{ger} = (\beta^{ger} + \eta_2) \times T_{it}^{ger} + \eta_1 \times T_{it}^{mat} + (\gamma^{ger} + \theta)X_{it} + u_i + \varepsilon_{it}^{ger} \tag{3-5}$$

---

[6] These models can be estimated by SUR to gain efficiency. However, efficiency gains are rather small in my setting and reduce standard errors at best by about five percent.

My final specification as seen in equation (4) and equation (5) can be regarded as a generalized first difference model across subjects[7] that allows for subject specific teacher trait effects β^{mat} and β^{ger}, as well as a distinct correlation of the individual fixed effect μ towards both subjects, measured through both η's.

In general, the WSBS estimation method accounts for the bias that would be induced by subject invariant student heterogeneity that is correlated with observable teacher characteristics or its latent factor of teacher effectiveness. It might for example be the case that more able students select themselves to better teachers. If higher teacher quality is correlated with one of the observed teacher characteristics T, this would yield to the omitted variable bias in the OLS estimation of β. As this is accounted for, my results are neither biased by selection of students to schools or selection to classrooms. As neither within-classroom nor courses tracking exist in 5th and 9th grade (the whole classroom is assigned to a single math or German teacher) this approach combined with the institutional setting is able to account for any type of selection to teachers, as long as the selection process is subject-invariant.

To identify a causal effect in this setting requires certain assumptions. These assumptions are less restrictive than a simple ordinary least squares regression[8] or regressions that solely account for school fixed effects. Yet, these assumptions are partially more restrictive than the standard VA approach. Although subject-invariant ability is accounted for, other potential confounding factors in the form of past inputs, unobserved current inputs and subject-specific ability might remain.

Unobserved subject invariant contemporary inputs are accounted for in the WSBS framework, while they are not in a standard VA model. In the case of value added specifications, unobserved current inputs in the form of subject-invariant inputs, and even subject varying inputs are not accounted for. Some VA estimates allow for a student fixed effects in gains, which is equivalent to my approach, besides, these fixed effects are subject-varying.

---

[7] Under the testable condition $\eta_1 = \eta_2$ my estimates would be equivalent to standard first difference or fixed effects estimators with subject specific parameter coefficients. Under the additional testable condition of $\beta^{mat} = \beta^{ger}$ the regression model simplifies to $\Delta y_{it} = \beta \Delta T_{it} + \Delta \varepsilon_{it}$, the simple first difference estimator across subjects.

[8] Each η measures the bias of which standard OLS would suffer due to the omission of unobserved student factors.

In the standard VA approach past educational inputs and time-invariant ability in a specific skill domain are, under certain assumptions, accounted for by the inclusion of the previous test score. However, as the WSBS approach does not control for the previous test score within each subject, one either has to make one of the following identifying assumptions; full decay of past educational inputs, full consideration through the subject-invariant fixed effect, conditional exogeneity on covariates, or zero correlation with current teacher traits. To control for past inputs and subject-specific ability heterogeneity and confine their potential of biasing the results, I use past grades in the respective subject as covariates (among the other above mentioned control variables). Grades are highly correlated with the test scores at hand ($\rho_5^{mat} =$ 0.53, $\rho_5^{ger} = 0.29$, $\rho_9^{mat} = 0.35$, $\rho_9^{ger} = 0.24$, within-classroom correlations are even higher with the exception of grade 5 mathematics) and therefore plausibly account for omitted variable bias due to selection to teachers through past subject-specific achievement.[9]

Both VA and WSBS estimates are potentially biased due to unobserved current inputs. However, as the NEPS data is very rich on variables, I claim that for this particular problem this study is dominating standard VA measures with few covariates.

Even if all confounding factors that are related to teachers are accounted for, my estimated coefficients have to be interpreted with caution. For the sake of simplicity consider the following educational production function that comprises perfectly adjusted test scores y of individual i such that their error component u is unrelated with teacher quality in terms of raising test scores Γ_{TS} of teacher j

$$\widetilde{y}_i = \tilde{\Gamma}_{TS,j} + \widetilde{u}_i \tag{3-a}$$

The most commonly employed specification for statistical testing of some trait T is a determinant of teacher quality is given by (3-b).

$$\tilde{\Gamma}_{TS,j} = \beta T_j + \tau_j \tag{3-b}$$

The whole teacher impact is divided into an observable and potentially correlated trait component T_j and unobserved teacher heterogeneity. Whether T_j is a

---

[9] Teachers following students over multiple years, potentially because of a high match quality, in a given subject are quite common in Germany which bears the potential of a bad control problem that can bias the results. Excluding grades does however not substantially alter the effects.

determinant or source of teacher quality cannot be answered in this framework, as it most likely correlated with error term of teacher quality, resulting in a classical omitted variable bias problem: $plim(\hat{\beta}) = \beta + \frac{cov(\tau_j, T_j)}{var(T_j)}$. Unless the trait has no effect at all ($\beta = 0$), or has an effect and that is conversely correlated with teacher quality, the direction of the coefficient is informative about the direction of the impact.

3.6   Results

This chapter presents the regression results from the models described in Chapter 3.5. All regressions include the covariates described in Table 1 and their means on the school level, as well as interactions with the respective starting cohort. All standard errors are clustered on the school level. It is important to note that teaching in mathematics and German differs substantially and that teachers that choose to become mathematics or German teachers differ in many observable characteristics, and likely in unobservable factors as well. Hence, it is crucial to analyze the two teacher populations separately and interpret the findings in that context. Hypotheses on the effect of the respective determinant are discussed and then reviewed in the context of the estimation results.

Experience has been found to be a determinant of teacher quality during the initial years of being in the profession in the US, thus reflecting transitional learning that flattens out after usually about three years. In the German setting, it may well be that no such learning exists, as student teachers practice teaching substantially during their training. As selection out of the profession is rare, years of experience might reveal interesting patterns in later years of the career in contrast to the US where selection out of job is rather common.

Figure 3-2 shows the results for both subject's teachers with experience categorized in three year intervals (Panel A) and five year intervals (Panel B). Interestingly, initial experience plays a positive role in mathematics, while it doesn't in German. Both subjects reveal a diminishing rate of experience hence turning the effect in math to zero and in German negative. The initial effect in math being similar to estimates from the US. However, regression coefficients and test statistics solely show borderline significant results.

Demographics are generally not regarded as main determinants of teacher quality; however they might reveal selection patterns into the profession. Fifty to eighty percent of teachers are female, depending on the subject taught and the cohort. On

the one hand, it may be that women are more successful at entering the teacher labor market and that this capability is correlated with teacher quality. On other hand, it may be that the outside option for able men is more attractive due to the gender wage gap in the private sector.

Table 3-3 presents results for basic teacher traits with gender in the first row. Columns (1) in both panels stem from simple OLS regressions including solely student covariates. Female teachers tend to perform worse than their male colleagues in both subjects with them being eleven percent of a SD less efficient in math and four percent in language. However, female teachers are disproportionately represented in lower secondary school tracks and accordingly the inclusion of a school track FE reduces the effect by half in math and surprisingly increases the effect to seven percent in German. There is evidence that selection into particular schools does not play the role it does in the US (Kristen 2003), hence one would not expect substantial changes due to the inclusion of a school FE in Column (3) which is indeed the case in both subjects. Including the student FE in Columns (4) yield the main results: Female teachers are eight percent of a standard deviation less efficient than their male counterparts with the null hypothesis of the subject specific coefficients being the same not rejected. Similar results are shown for migrant teachers in the third line. Age is generally not thought of as a determinant of student achievement and furthermore highly correlated with experience in Germany, hence, one would expect similar results as for experience. However, age is measured more noisily in the data with ten year intervals and with experience revealing the inverse U-shape in math one would expect a zero effect in mathematics which is indeed the case. The effect of experience being solely negative in German, the linear regression coefficients reveal the negative slope over age that however vanishes once the student FE is included.

Selection into the profession that cannot be measured with regular administrative teacher data is obviously an important determinant of teacher quality (Nagler, Piopiunik, and West, 2015). However, NEPS' rich data allows looking into the aspects of career choice a teacher used to consider before joining the workforce. Enjoying teaching or the subject's content and job security that are not presented in Table 3-4 have no predictive power for teacher quality, quite the opposite: Estimates are rather precisely measured to be around zero.

Further training is shown to affect workers' productivity; however estimates depend largely on the particular setting and estimation (Zwick 2006). Additionally, the

specific types chosen by teachers are non-random and significant findings might represent again selection rather than the effectivity of a practice. All estimates of training practices presented in Table 3-5 in Columns (3) and (4) show non-significant results that are measured more noisily than results from Table 3-4. If anything, qualification programs and working groups may have potential to increase teacher quality.

Teaching practices are found to affect student learning. However, these practices are likely to be highly endogenous to a classroom's composition. Hence, it might be more promising to look into teaching philosophy's or a teacher's general educational goals that are unaffected by students. As I can additionally account for selection of students to teachers, these estimates may provide insightful results in Table 3-6 and Table 3-7. Indeed, and in line with literature on practices, the effects of teaching facts and quietness in classroom in Table 3-6 are positive and statistically significant predictors of student achievement, which may be interpreted as the positive effect of "traditional teaching" (Schwerdt, and Wuppermann, 2011). Besides subject knowledge and problem solving, all goals presented in Table 3-7 may yield negative estimates as the share of time committed to teaching subject knowledge that is tested in math and German is less in a classroom taught by a teacher whose focus is on other things. However, that does not seem to be the case with all coefficients in both subjects being equal in Columns (4) and close to zero. Student's skills like self-confidence are found to be positive correlates of their achievement; however the discussed effects may cancel each other out.

Non-cognitive skills are found to have strong predictive power for labor market outcomes (Carneiro, Crawford, and Goodman, 2007), it is, however, unclear how and in what combination a teacher's skills are beneficial for students. The interaction of colleagues is likely to be a manifestation of these skills at work, hence it is surprising to find rather precise zero effects in Table 3-8. The exception being teaching in joint lessons with another colleague, which is measured on a 5 point scale from rarely to frequently: An increase on that scale of one is associated with a decrease of teaching quality by 3.4 percent of a SD. That practice's negative effect may be driven by the doubling of the class size that cannot be compensated by the doubling of the teaching staff.

The reasons of perceived stress may also shed some light on the channels at work of teacher quality. Results that are not presented here include stress because of lack of time for curriculum, lack of career opportunities, and lack of recognition as well as

rivalry between colleagues that do not yield significant findings. As do most of the presented findings in Table 3-9 with the exception being stress due to insecurities about teaching methods in mathematics. More stress about these decreases student achievement gains by 4.5% of a SD, which is very much in line with the results for teaching philosophies.

Finally, Table 3-10 provides information of the impact of teachers' previous labor market experience. Teachers might acquire skills on the labor market that may be beneficial. However, having previous pedagogical experience, having participated in the military or civil service or any other or vocational training is not predictive of teacher effectiveness in any of the specifications in either subject.

In general, the results provide some suggestive evidence that selection of students and teachers in Germany is rather low. Solely including school type FEs (Columns (2)) is usually enough to yield similar results as individual FE models (Columns (4)). Once selection into a particular school (Columns (3)) is accounted, results do not substantially differ in any of the results. Furthermore, adjusted R^2 show that controls and school type fixed effects can account for about forty percent of the variation. Including a school FE increases the R^2 to solely forty-eight.

3.7   Conclusion

This study analyzes a large set of potential determinants of teacher quality in Germany. The contribution of the study is twofold: Firstly, analyzing a large set of potential predictors can spark future research into certain predictors. Secondly, as data in the US is usually restricted to administrative data, it sheds light on potential determinants that could not be investigated previously.

Experience of mathematics teachers reveals the same increasing pattern over time in occupation, but not so in German. Both subjects show a diminishing rate of experience that makes gains of experience associated with losses in effectiveness over time, potentially due to a loss of intrinsic motivation over time. Teachers with a migration background and women tend to worse than their counterparts. Quietness in the classroom and teaching facts as being of importance for teachers is associated with higher student test score gains.

Future research will look into the mediating channels of the found effects in experience, gender, migration background, teaching philosophies and holding joint

lessons with another teacher. Findings in these areas may later be used to design policies that foster selection in or out of the profession based on these traits.

Figure 3-1: Tracking in Germany until grade 10
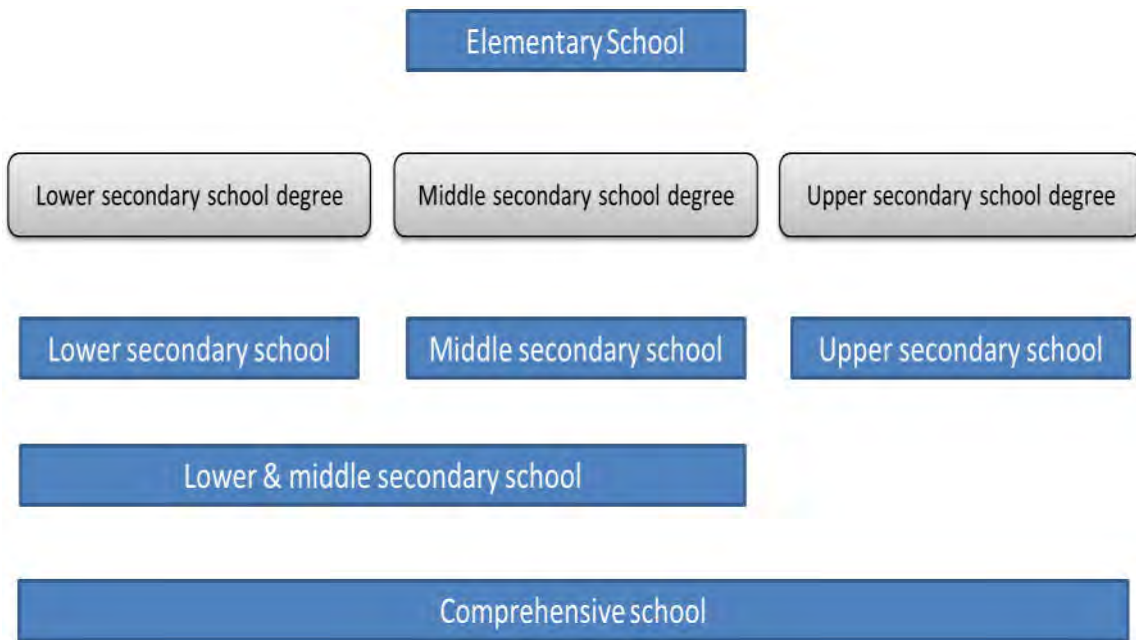
Figure 3-2: Teacher effectiveness over time in occupation by subject
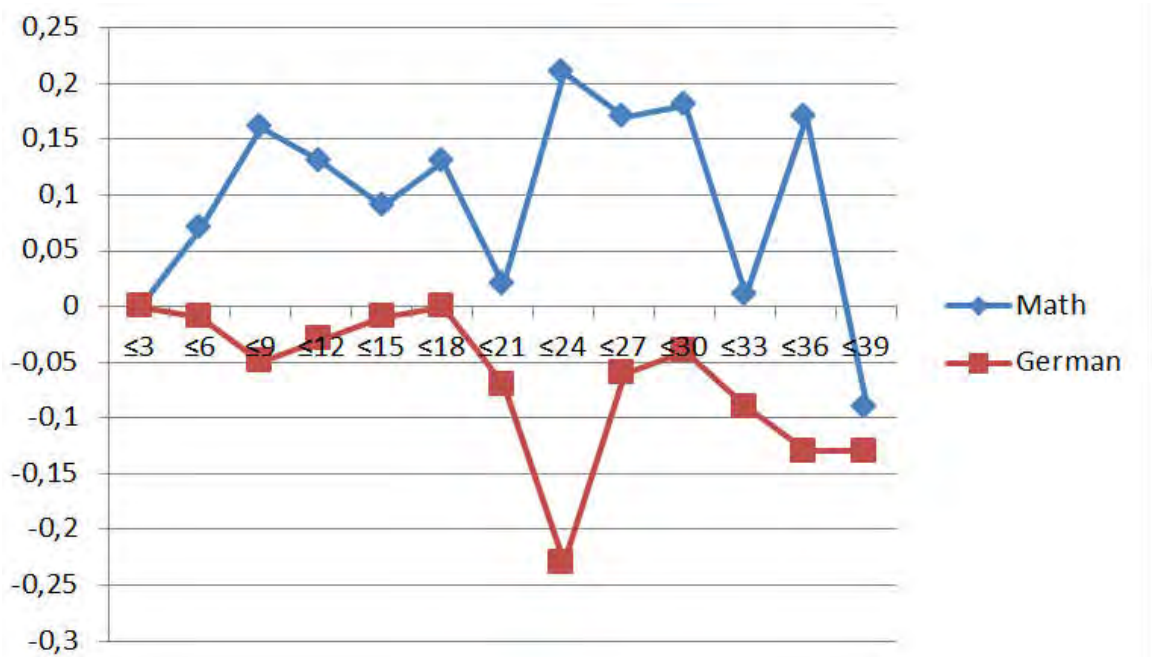
A: Experience in three year intervals
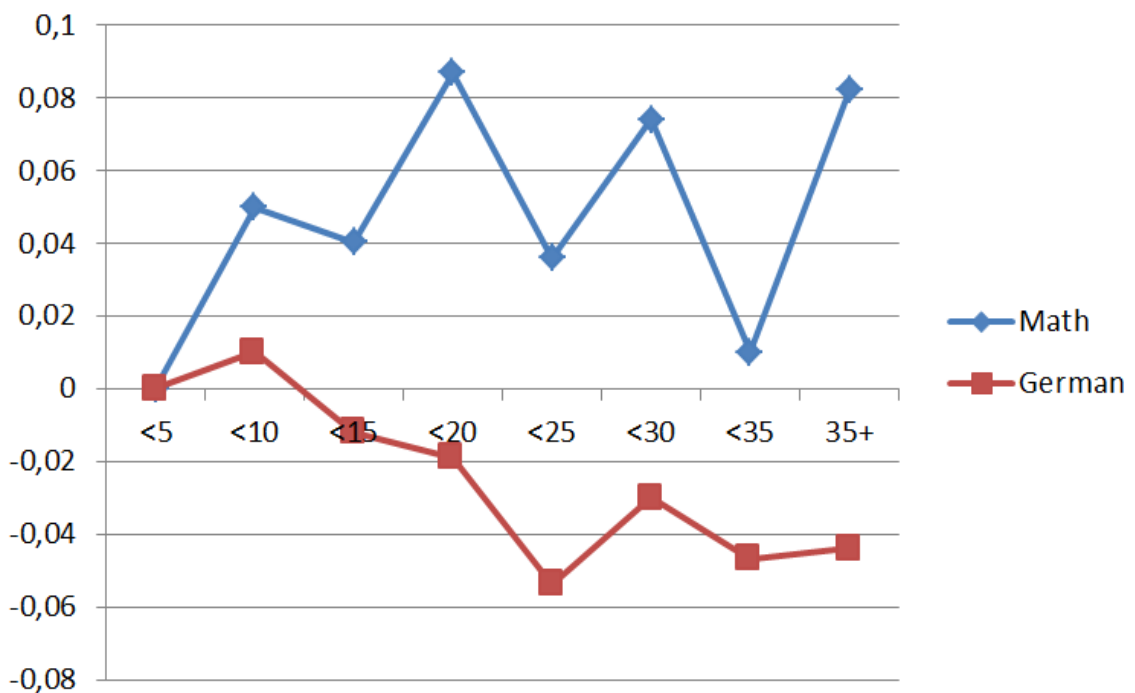
B: Experience in five year intervals

Table 3-1: Descriptive statistics of older cohort's students by type of school

	Full	CO	LO	ML	MI	HI
Female	0.49	0.49	0.43	0.51	0.48	0.53
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age	15.02	15.01	15.27	15.18	15.01	14.79
	(0.63)	(0.59)	(0.69)	(0.68)	(0.61)	(0.50)
Born abroad	0.06	0.07	0.11	0.05	0.05	0.03
	(0.24)	(0.26)	(0.31)	(0.22)	(0.22)	(0.18)
Household size	4.65	4.61	4.92	4.49	4.66	4.47
	(1.84)	(1.84)	(2.04)	(1.99)	(1.76)	(1.61)
Mother born abroad	0.20	0.24	0.30	0.10	0.18	0.14
	(0.40)	(0.43)	(0.46)	(0.30)	(0.38)	(0.35)
Father born abroad	0.21	0.24	0.32	0.10	0.19	0.14
	(0.40)	(0.43)	(0.47)	(0.31)	(0.39)	(0.35)
PC availability	1.29	1.27	1.33	1.31	1.28	1.26
	(0.47)	(0.46)	(0.50)	(0.49)	(0.46)	(0.45)
Retention	0.20	0.17	0.32	0.26	0.21	0.08
	(0.40)	(0.37)	(0.47)	(0.44)	(0.41)	(0.27)
Number of books	3.74	3.73	2.97	3.29	3.71	4.54
	(1.49)	(1.52)	(1.42)	(1.44)	(1.34)	(1.25)
Perceptual speed score	58.65	59.32	56.19	58.40	58.88	60.39
	(13.79)	(14.07)	(15.62)	(13.23)	(13.69)	(11.95)
Reasoning score	8.66	8.68	7.16	7.98	8.87	9.90
	(2.46)	(2.35)	(2.58)	(2.48)	(2.15)	(1.77)
German: Grade Point	2.84	2.64	3.06	2.84	2.95	2.63
	(0.86)	(1.13)	(0.80)	(0.79)	(0.77)	(0.82)
Math: Grade Point	2.94	2.76	3.10	3.02	2.99	2.80
	(1.06)	(1.30)	(1.04)	(0.95)	(1.06)	(1.02)
German: Test score	0.00	-0.11	-0.49	-0.38	0.11	0.48
	(1.00)	(0.98)	(0.93)	(0.96)	(0.91)	(0.88)
Math: Test score	0.00	-0.37	-0.61	-0.48	-0.07	0.81
	(1.00)	(0.80)	(0.70)	(0.77)	(0.71)	(0.95)
Observations	7273	687	2038	647	1570	2331

Means of variables with standard deviations in parentheses for the full sample in the first column. Each following column represents a particular school type. LO indicates the lowest secondary school track, ML secondary schools that mix middle and lowest tracks, MI the middle secondary school track, CO comprehensive schools and HI the highest secondary school track.

Table 3-2: Basic descriptive statistics of teachers by cohort and subject taught

	(1) Younger cohort						(2) Older cohort					
	Full	Min	Max	Math	German	Diff	Full	Min	Max	Math	German	Diff
Female	0.72 (0.45)	0.0	1.0	0.68 (0.47)	0.77 (0.42)	-0.24 (0.42)	0.56 (0.50)	0.0	1.0	0.48 (0.50)	0.64 (0.48)	-0.27 (0.42)
Born before 1950	0.06 (0.24)	0.0	1.0	0.07 (0.26)	0.05 (0.23)	-0.04 (0.21)	0.09 (0.29)	0.0	1.0	0.09 (0.29)	0.09 (0.28)	-0.10 (0.27)
Born 1950-1959	0.35 (0.48)	0.0	1.0	0.37 (0.48)	0.33 (0.47)	-0.12 (0.32)	0.39 (0.49)	0.0	1.0	0.41 (0.49)	0.36 (0.48)	-0.20 (0.37)
Born 1960-1969	0.18 (0.39)	0.0	1.0	0.18 (0.39)	0.18 (0.39)	-0.09 (0.29)	0.23 (0.42)	0.0	1.0	0.23 (0.42)	0.22 (0.41)	-0.16 (0.35)
Born 1970-1979	0.24 (0.43)	0.0	1.0	0.22 (0.42)	0.26 (0.44)	-0.16 (0.36)	0.20 (0.40)	0.0	1.0	0.18 (0.38)	0.21 (0.41)	-0.16 (0.35)
Born after 1979	0.16 (0.37)	0.0	1.0	0.16 (0.36)	0.17 (0.38)	-0.11 (0.31)	0.10 (0.30)	0.0	1.0	0.09 (0.28)	0.12 (0.32)	-0.09 (0.28)
Time in profession	4.23 (2.55)	1.0	9.0	4.49 (2.58)	3.98 (2.49)	-0.66 (1.52)	4.57 (2.48)	1.0	8.0	4.72 (2.45)	4.41 (2.50)	-1.01 (1.66)
Time at school	2.76 (2.06)	1.0	8.0	2.96 (2.09)	2.56 (2.02)	-0.64 (1.56)	3.20 (2.21)	1.0	8.0	3.31 (2.27)	3.09 (2.16)	-0.89 (1.52)
Migration status	0.09 (0.40)	0.0	2.0	0.10 (0.41)	0.08 (0.39)	-0.06 (0.33)	0.06 (0.31)	0.0	2.0	0.06 (0.29)	0.06 (0.32)	-0.07 (0.33)
Other field of study	0.20 (0.40)	0.0	1.0	0.19 (0.39)	0.20 (0.40)	-0.13 (0.34)	0.20 (0.40)	0.0	1.0	0.21 (0.41)	0.18 (0.39)	-0.12 (0.30)
HS GPA	2.72 (0.60)	1.4	4.0	2.79 (0.61)	2.65 (0.58)	-0.13 (0.33)	2.60 (0.57)	1.2	4.0	2.65 (0.58)	2.56 (0.55)	-0.17 (0.32)
1SE GP	2.87 (0.55)	1.5	4.0	2.81 (0.54)	2.94 (0.56)	-0.27 (0.43)	2.81 (0.57)	1.0	4.0	2.77 (0.56)	2.86 (0.57)	-0.28 (0.43)
2SE GP	2.91 (0.61)	1.0	4.0	2.86 (0.62)	2.97 (0.60)	-0.23 (0.43)	2.87 (0.63)	1.0	4.0	2.86 (0.62)	2.88 (0.64)	-0.24 (0.42)
Observations	371			185	186	185	1058			536	522	522

Means of variables with standard deviations in parentheses and minima and maxima for the full sample by each starting cohort in the first column of each panel. The two following column show means and standard deviations in parantheses by subject taught. The last column shows the means of the differences between teacher traits within the classroom and its standard deviation.

Table 3-3: Estimation results for teachers' basic traits

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Female	-0.108	-0.043	-0.025	-0.076	-0.040	-0.072	-0.088	-0.076
N=7692 / 7691 / 7606	(0.044)	(0.028)	(0.041)	[0.081]	(0.046)	(0.038)	(0.063)	[0.081]
Age	-0.018	0.010	0.004	0.005	-0.029	-0.007	-0.033	0.005
N=7728 / 7714 / 7643	(0.010)	(0.006)	(0.008)	[0.565]	(0.009)	(0.008)	(0.014)	[0.565]
Migration background	-0.072	-0.011	-0.014	-0.097	-0.036	0.036	-0.129	-0.097
N=7726 / 7708 / 7630	(0.079)	(0.039)	(0.038)	[0.068]	(0.100)	(0.073)	(0.136)	[0.068]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-4: Estimation results for teachers' aspects of career choice

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Family compatibility N=7663 / 7695 / 7571	0.042 (0.029)	-0.005 (0.019)	-0.001 (0.026)	0.034 [0.301]	0.007 (0.024)	0.020 (0.020)	0.044 (0.047)	0.034 [0.301]
Contact with people N=7711 / 7781 / 7688	-0.055 (0.038)	0.011 (0.023)	-0.006 (0.034)	-0.003 [0.937]	-0.012 (0.037)	-0.048 (0.030)	-0.007 (0.046)	-0.003 [0.937]
Much spare time N=7447 / 7579 / 7263	0.029 (0.030)	0.005 (0.019)	0.045 (0.025)	0.032 [0.273]	-0.040 (0.027)	0.017 (0.023)	0.003 (0.040)	0.032 [0.273]
Good pay N=7616 / 7673 / 7512	0.020 (0.033)	0.022 (0.023)	0.038 (0.036)	0.035 [0.372]	0.014 (0.029)	0.045 (0.025)	0.066 (0.044)	0.035 [0.372]
Challenge N=7653 / 7751 / 7608	-0.082 (0.032)	-0.063 (0.019)	-0.052 (0.028)	0.003 [0.941]	0.019 (0.031)	-0.039 (0.027)	-0.010 (0.044)	0.003 [0.941]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-5: Estimation results for teachers' participation in further training

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Courses / workshops N=7647 / 7602 / 7506	-0.108 (0.049)	-0.046 (0.031)	-0.004 (0.039)	-0.050 [0.139]	-0.064 (0.049)	-0.052 (0.044)	-0.119 (0.068)	-0.050 [0.139]
Conferences / seminars N=7576 / 7462 / 7304	-0.005 (0.046)	-0.033 (0.029)	-0.034 (0.038)	-0.074 [0.104]	-0.030 (0.043)	-0.030 (0.036)	0.003 (0.071)	-0.074 [0.104]
Qualification programs N=7534 / 7471 / 7271	0.011 (0.067)	0.000 (0.043)	0.069 (0.064)	0.120 [0.211]	0.020 (0.076)	0.020 (0.072)	0.001 (0.135)	0.120 [0.211]
Visiting other classrooms N=7533 / 7487 / 7293	0.023 (0.061)	0.023 (0.034)	0.002 (0.055)	0.006 [0.810]	0.053 (0.055)	0.013 (0.046)	0.047 (0.067)	0.006 [0.810]
Working group N=7556 / 7483 / 7319	0.033 (0.050)	-0.062 (0.030)	0.030 (0.044)	0.078 [0.205]	0.026 (0.044)	-0.002 (0.037)	0.082 (0.076)	0.078 [0.205]
Research work N=7540 / 7519 / 7444	-0.107 (0.080)	-0.113 (0.044)	-0.037 (0.063)	-0.029 [0.750]	-0.041 (0.057)	-0.058 (0.047)	0.110 (0.078)	-0.029 [0.750]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-6: Estimation results for teachers' teaching philosophy

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Students making own decisions	0.097	0.074	0.032	0.010	-0.007	0.019	0.011	0.010
N=7664 / 7650 / 7532	(0.034)	(0.020)	(0.029)	[0.789]	(0.034)	(0.031)	(0.044)	[0.789]
Easing students exploring	-0.060	-0.014	-0.014	-0.005	-0.054	-0.006	-0.023	-0.005
N=7703 / 7632 / 7565	(0.038)	(0.023)	(0.035)	[0.900]	(0.033)	(0.030)	(0.043)	[0.900]
Learning by doing	-0.060	-0.043	-0.064	-0.061	-0.028	-0.036	-0.067	-0.061
N=7731 / 7705 / 7653	(0.038)	(0.023)	(0.030)	[0.111]	(0.034)	(0.029)	(0.045)	[0.111]
Teaching facts	0.013	0.011	0.019	0.056	-0.010	-0.022	-0.015	0.056
N=7702 / 7624 / 7593	(0.031)	(0.019)	(0.027)	[0.046]	(0.029)	(0.024)	(0.052)	[0.046]
Quietness in classroom	-0.009	0.015	0.063	0.132	-0.061	-0.045	-0.116	-0.051
N=7735 / 7692 / 7319	(0.037)	(0.024)	(0.037)	[0.006]	(0.040)	(0.034)	(0.055)	[0.457]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ² test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-7: Estimation results for teachers' goals for students' education

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Subject knowledge	0.157	0.041	0.024	0.018	0.073	0.018	0.016	0.018
N=7684 / 7716 / 7608	(0.036)	(0.024)	(0.034)	[0.652]	(0.032)	(0.028)	(0.059)	[0.652]
Problem solving	0.154	0.013	0.008	0.013	0.060	0.000	-0.063	0.013
N=7711 / 7754 / 7672	(0.040)	(0.028)	(0.039)	[0.792]	(0.039)	(0.034)	(0.060)	[0.792]
Knowledge for later life	-0.157	0.007	0.002	0.011	-0.053	0.005	-0.035	0.011
N=7662 / 7673 / 7599	(0.027)	(0.018)	(0.023)	[0.717]	(0.025)	(0.021)	(0.031)	[0.717]
Self-confidence	-0.191	-0.052	-0.004	0.025	-0.056	0.011	-0.031	0.025
N=7702 / 7662 / 7598	(0.048)	(0.027)	(0.035)	[0.601]	(0.040)	(0.037)	(0.069)	[0.601]
Social skills	-0.196	-0.028	0.013	0.023	-0.027	0.042	-0.005	0.023
N=7732 / 7729 / 7686	(0.051)	(0.030)	(0.045)	[0.616]	(0.044)	(0.042)	(0.066)	[0.616]
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-8: Estimation results for teachers' interaction with colleagues

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Sharing teaching materials	-0.063	-0.006	0.006	-0.018	-0.047	-0.008	-0.005	-0.018
N=7713 / 7723 / 7635	(0.017)	(0.011)	(0.016)	[0.392]	(0.017)	(0.014)	(0.023)	[0.392]
Planning lessons	-0.071	-0.008	-0.024	-0.021	-0.059	-0.015	0.006	-0.021
N=7721 / 7673 / 7593	(0.015)	(0.010)	(0.018)	[0.333]	(0.015)	(0.014)	(0.025)	[0.333]
Discussing students' progress	-0.031	0.007	0.033	0.013	-0.065	-0.039	-0.005	0.013
N=7715 / 7715 / 7271	(0.020)	(0.012)	(0.019)	[0.501]	(0.019)	(0.015)	(0.024)	[0.501]
Joint lessons	-0.059	-0.004	0.002	-0.034	-0.066	-0.026	-0.009	-0.034
N=7687 / 7725 / 7611	(0.015)	(0.009)	(0.014)	[0.040]	(0.013)	(0.012)	(0.019)	[0.040]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-9: Estimation results for teachers' reasons of stress in profession

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Students' differential capabilities	-0.018	0.007	0.025	-0.001	-0.065	-0.044	-0.009	-0.001
N=7733 / 7778 / 7506	(0.022)	(0.015)	(0.022)	[0.968]	(0.021)	(0.019)	(0.030)	[0.968]
Teaching methods	-0.014	0.010	0.058	0.045	-0.062	-0.028	-0.010	-0.022
N=7694 / 7786 / 7677	(0.026)	(0.015)	(0.024)	[0.102]	(0.023)	(0.020)	(0.029)	[0.509]
Students' behavior	-0.031	0.002	0.036	0.015	-0.043	-0.008	-0.005	0.015
N=7746 / 7771 / 7729	(0.023)	(0.014)	(0.020)	[0.542]	(0.021)	(0.018)	(0.030)	[0.542]
Lesson planning	-0.013	-0.002	-0.013	-0.020	-0.010	-0.026	-0.009	-0.020
N=7701 / 7785 / 7683	(0.022)	(0.015)	(0.022)	[0.307]	(0.022)	(0.020)	(0.035)	[0.307]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ² test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

Table 3-10: Estimation results for teachers' careers before becoming a teacher

	Math				German			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Pedagogical experience N=7757 / 7750 / 7703	-0.190 (0.053)	-0.031 (0.035)	-0.074 (0.047)	-0.021 [0.673]	-0.006 (0.044)	0.038 (0.038)	0.057 (0.060)	0.106 [0.112]
Military / civil service N=7094 / 7177 / 6664	-0.011 (0.052)	0.010 (0.033)	0.010 (0.039)	0.053 [0.236]	-0.023 (0.044)	-0.003 (0.037)	0.055 (0.057)	0.053 [0.236]
Other job N=7534 / 7471 / 7271	-0.031 (0.058)	-0.023 (0.035)	-0.022 (0.052)	-0.052 [0.438]	0.022 (0.051)	-0.023 (0.040)	-0.021 (0.071)	-0.052 [0.438]
Vocational training N=7533 / 7487 / 7293	0.169 (0.071)	0.059 (0.046)	0.056 (0.063)	-0.490 [0.459]	0.016 (0.058)	-0.021 (0.047)	0.006 (0.096)	-0.490 [0.459]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School type FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School FE	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes

Dependent variables: Test score in mathematics and German. Left panel shows results for mathematics, right panel for German. Columns (1) through (3) are estimated separately by OLS with cluster robust standard errors on the school level in parentheses. Columns (4) shows results from a joint correlated random effects SUR-estimation with p-values of a χ^2 test of the coefficient being zero in brackets. Below each variable name are the estimation sample sizes for math, German and the joint estimation respectively. No superscript in columns (4) means that the hypothesis test of equality of η could not be rejected and that these regression coefficients were restricted to equality. If β estimates in columns (4) are equal, the hypothesis of them being equal could not be rejected and they were as well restricted to equality.

3.8 Appendix

Figure 3-3: Student observation loss over the merging procedure in the younger cohort

1) No special education: 5,525 // 100%

2) Test score data: 5,503

3) Student questionnaire: 5,187

4) German teacher ID: 2,829

5) Math teacher ID: 2,234; K=185

Final Sample: N=2,234 // 40.4%

Schools:

$K = 131$
$N_K(min) = 1$
$N_K(max) = 66$
$N_K(mean) = 17.05$

1 Classroom:
$N = 867$
$K = 77$
$N_K(min) = 1$
$N_K(max) = 23$
$N_K(mean) = 11.26$

2 Classrooms:
$N = 1,367$
$K = 112$
$N_K(min) = 1$
$N_K(max) = 34$
$N_K(mean) = 12.21$
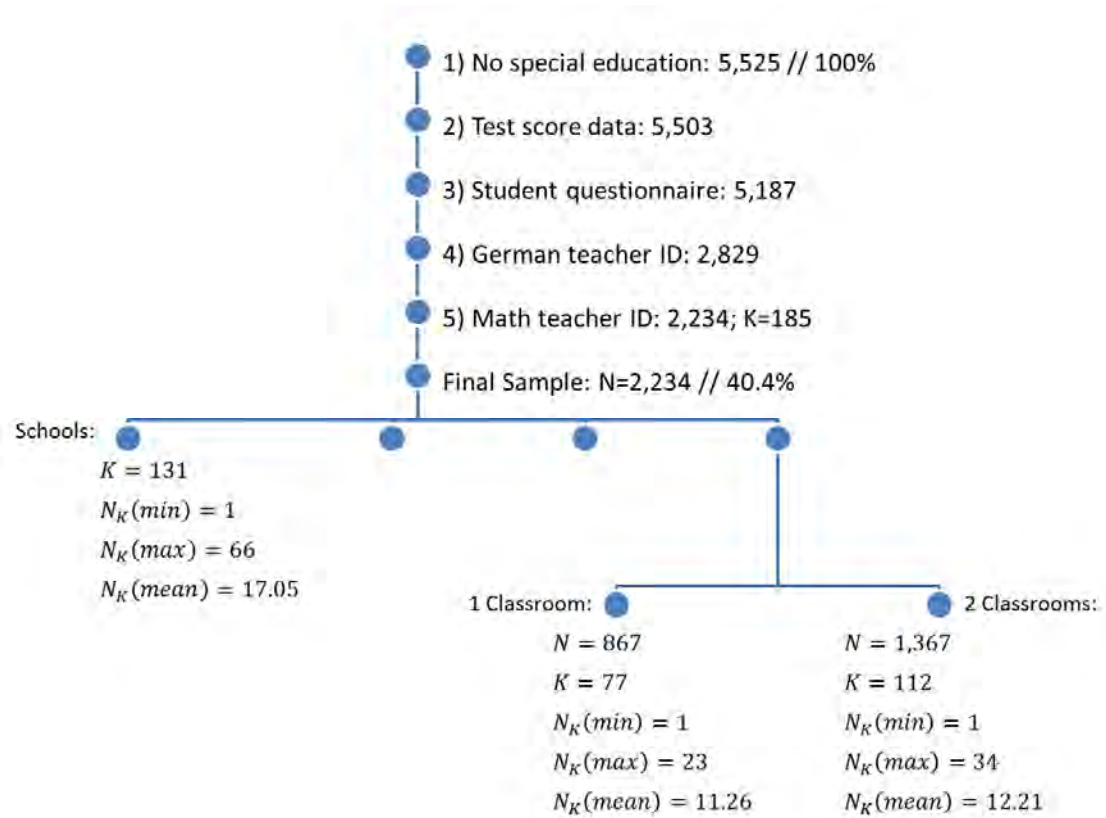
Figure 3-4: Student observation loss over the merging procedure in the older cohort

1) No special education: 15,239 // 100%

2) Test score data: 15,017

3) Student questionnaire: 14,490

4) German teacher ID: 9,587

5) Math teacher ID: 7,273; K=534

Final Sample: N=7,273 // 47.7%

Schools:

$K = 358$
$N_K(min) = 2$
$N_K(max) = 61$
$N_K(mean) = 20.32$

1 Classroom:

$N = 2,839$
$K = 198$
$N_K(min) = 2$
$N_K(max) = 30$
$N_K(mean) = 14.33$

2 Classrooms:

$N = 4,434$
$K = 320$
$N_K(min) = 1$
$N_K(max) = 33$
$N_K(mean) = 13.86$
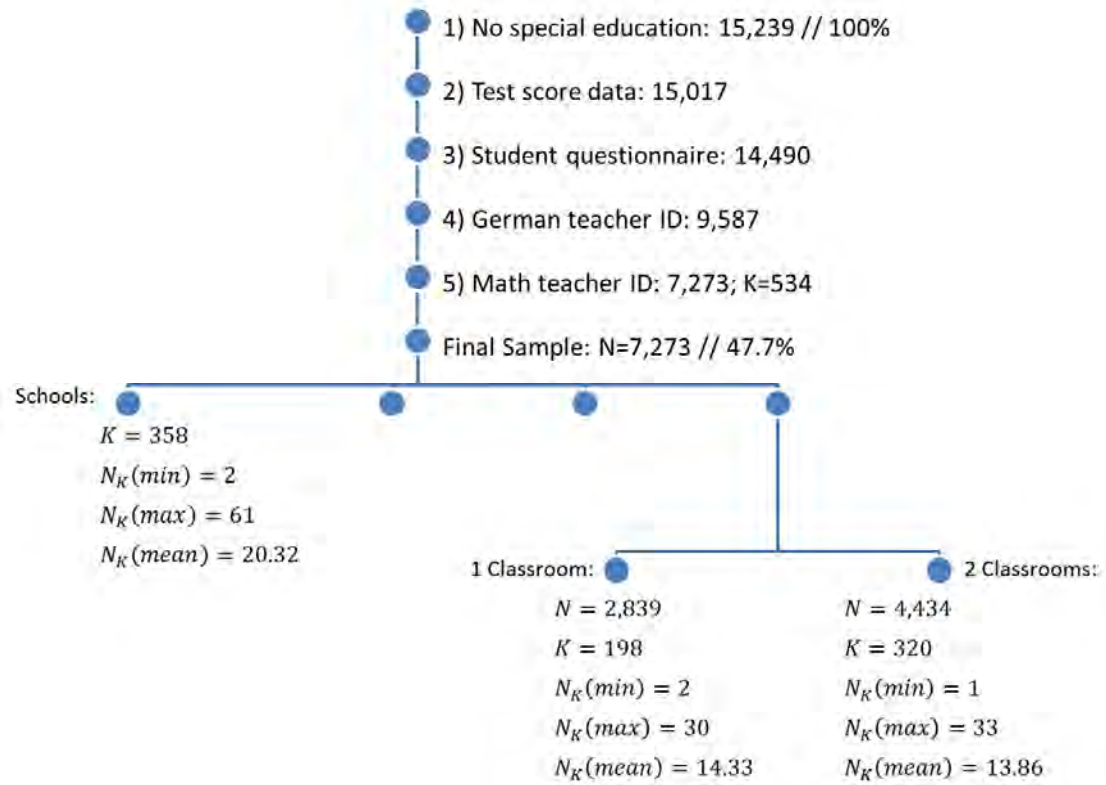
4 Gender differentials in test scores and teacher assessments in Germany

4.1 Introduction

Are there systematic differences in the way teachers grade their male and female students conditional on the same performance? Experimental[1] and observational[2] studies have shown that boys and children with a migration background tend to be graded worse conditional on the same performance.

Investigating school grade differentials conditional on the same performance is important for various reasons. First, Altonji, and Pierret (2001) have shown that high school grades are highly correlated with wages at labor market entry. Hence, systematic differences in grading schemes that are not caused by actual performance differences may induce wages that are not reflecting productivity discrepancies but factors that an employer might not want to take into account at the employment decision. These avoidable uncertainties might induce inefficiencies. Second, systematic grading differences by a student's gender may explain the gender role reversal in education over the past decade, as Goldin, Katz, and Kuziemko (2006) show that there exist advantages for females in the US school environment.

4th and 5th class grades can be even more consequential than the ones from higher school years in a tracked school system. In most German states, grades in these years determine the secondary school track and, thereby, future career paths as only the

---

[1] e.g. (Hanna, and Linden, 2009; Hinnerich, Höglin, and Johannesson, 2011a, 2011b; Sprietsma, 2012)

[2] e.g. (Burgess, and Greaves, 2013; Cornwell, Mustard, and Van Parys, 2013; Lavy, 2008)

highest secondary school track provides direct, unrestricted access to tertiary education.

Using the rich data set of NEPS, I investigate the relationship of 5th and 6th class students' genders and their previous year's grades in math, science and German. NEPS consists of extensive questionnaires for students, parents, teachers and school principals that allow me to control for many determining factors of grades. Besides grades and basic characteristics (e.g. gender, age, migration status and socio-economic status), it includes information about life satisfaction, intelligence, leisure time activities, grades and non-cognitive skills. Most importantly, it includes objective measures of performances in math, science and German.

Using fixed effects estimators to account for unobserved heterogeneities, I find indications of subject specific grading by a student's gender. While girls are, conditional on all controls and a classroom fixed effect (FE), advantaged by 22.2% of a standard deviation (SD) in German, they are disadvantaged by 19.7% of a SD in math relative to boys. No significant gender gap exists for science. These findings are robust to many different specifications. Investigating whether the gender gap can be explained by heterogeneous teacher effects, I find no altering effects by teacher characteristics (e.g. migrant status, gender) or different in-class time use.

The remainder of this part is structured as follows: the following chapter presents the data. Chapter 4.3 describes the estimation strategy and presents headline results and their discussion. Chapter 4.4 concludes.

4.2   Data

To identify and explain gender grade differentials, I use data from the 2010 and 2011 waves of NEPS on a cohort that was first tracked in 5th class. NEPS data was collected via a stratified sampling procedure: At first, a random sample of schools was drawn. Within those schools, up to two random classes were selected to participate.

Therefore, the data used in this study consists of student observations in 5th and 6th class. The students were tested in mathematics, science and German and were also asked about their last final grades in these subjects. Figure 4-1 illustrates the timing and availability of the respective data in each class for this cohort.

To investigate the relationship between a student's gender and his or her grades while conditioning on performance, I use pairs of last final grades and test scores

from the beginning of the subsequent year. As grade and test score are solely divided by school holidays, they are based on the same underlying performance. One of these pairs is available for math and science, two for German. The only pair that can be linked to the respective teacher is the pair of 5th class final grade in German and the 6th class test score.

I limit the analysis to students who were taught by one teacher in each subject and to those that did not require any form of special education. Apart from testing, students, parents, principals, as well as German and math teachers were extensively asked about background information, which allows me to control for many other determining factors of grades. Participation for each of these individuals was voluntary and about 5% of the students did not participate in the testing.

All grades and test scores are standardized to have a mean of zero and a standard deviation of one with higher values indicating better performance. Table 4-1 presents some first descriptive evidence, showing firstly, that boys are on average better in all standardized tests while they are only graded favorably in mathematics. Secondly, boys tend to show less beneficial social behavior compared to girls.

4.3 Empirical framework

To investigate the relationship between grades, test scores and gender, I follow the approach of Cornell, Mustard and Van Parys (2013) and model the grade production function in a more general form:

$$grade_i^k = \alpha^k \times female_i + \gamma^k \times testscore_i^k + \lambda_i^k X + \mu_i^k + \tau_j^k + \xi_m^{k} + \varepsilon_i^k \qquad (4\text{-}1)$$

The grade of student i in subject k is a linear, additive function of a gender indicator, the respective test score and control variables X with their subject specific coefficients α , γ and λ. The unobserved factors comprise a teacher τ, school ξ, individual μ and idiosyncratic ε component. To account for unobserved teacher and school effects I use classroom and school fixed-effects estimators. However, as it is impossible to account for unobserved student heterogeneity with this data, I will use a large set of control variables to minimize omitted variable bias. Table (2) presents the main estimation results.

Only using within school variation and conditioning on standardized test scores in mathematics, science and German, estimation results from setting (1) show that girls are graded less favorably than their male counterparts in mathematics while the opposite is true for German. No gender difference is revealed for science

through all settings. These first results may be driven by unobserved teacher factors, but including a classroom fixed effect in specification (2) does not substantially change the estimates. Additionally controlling for student background characteristics in setting (3) does not alter the results noticeably either.

Setting (4) presents the headline results and also controls for non-cognitive skills as in Cornwell, Mustard and Van Parys (2013). In contrast to their findings, the addition of non-cognitive skills does not explain the gender-grade gap. The set of variables measuring non-cognitive skills includes the students' results in two SDQ questionnaires These non-cognitive skill measures plausibly add to the explained variation and take away explanatory power of the test scores in grades, as non-cognitive skills are important for classroom activities weakly correlated with test score performance. However, the magnitude and direction of the gender effect remains the same.

4.4 Results

These results may still be driven by omitted student variables. A first-difference approach across subjects would account for this unobserved student heterogeneity and, thus, yield unbiased results due to the omission of student variables, but both coefficients could not be individually identified anymore. However, their difference, $\Delta\alpha$, still is: Conducting this FD approach as a robustness check yields no statistically different results from the difference of the two gender coefficients in specification (4), suggestive evidence for the robustness of this finding.

However, there are a few remaining potential threats to this identification strategy. First, it might be the case that male and female students participate differently in the classroom conditional on the same test score. If classroom participation is determined by performance, but not as a simple linear function of the respective test score, many potential bias scenarios are imaginable. Consider the case in which male students, no matter their actual performance, do not participate in the German classroom, while female students do according to their performance. Relative to girls, boys would get worse grades conditional on the same test score, as classroom participation is an important determinant of grades. The estimator related to the female indicator variable would, therefore, be confounded by different classroom participation patterns. Checking for this by including gender - test score interactions does not reveal any significant gender specific test score effects.

Secondly, it may be the case that teachers' grading patterns confound these estimates. As well as in the classroom participation case, many potential biases may arise. Consider for example the case in which an average teacher in German grades on a curve and girls outperform boys. Boys are therefore, holding everything else constant, relatively pushed down the grade distribution, although the underlying performance gap might not suggest so. The female effect in a regression conditioning on test scores would therefore be overestimated. Running regressions on the small subset of students who are taught by the same teacher, thus implicitly assuming that a teacher would use the same grading scheme in both subjects, shows that coefficients remain at the same magnitude, although they are not significantly different from zero anymore due to the small sample size.

Further using teacher data with the German test score - grade combination in 5th class, I cannot find substantially altering effects for the gender estimator. These analyses include differences by teachers' basic traits like age, gender and origin (East and West Germany), as well as teachers' self-reported determinants of final grades in the form of classroom participation, essay writing, dictation, written tests or homework assignments.

4.5 Conclusion

Conducting an analysis of grade determinants, I find that gender plays a crucial role in grade production. Accounting for several potential identification threats and testing various specifications to explain the gender gap, I find no factor that can do so.

As these results are not driven by unobserved teacher traits and - due to the large set of control variables - may not be by unobserved student heterogeneity, one could, if omitted student variable bias is truly accounted for, interpret them as quasi-causal effects: Solely based on his or her gender, a student might be assessed differently for example through gender stereotype grading of teachers. However, further research is necessary to support this claim.

Future research should a) examine the influence of student-teacher interactions on grade production more thoroughly, b) find ways to account for unobserved student heterogeneity while keeping the gender coefficient identifiable and c) supplementary investigate other potential grading gaps (e.g. migrant status).

Figure 4-1: Data availability and timeline of testing and grading
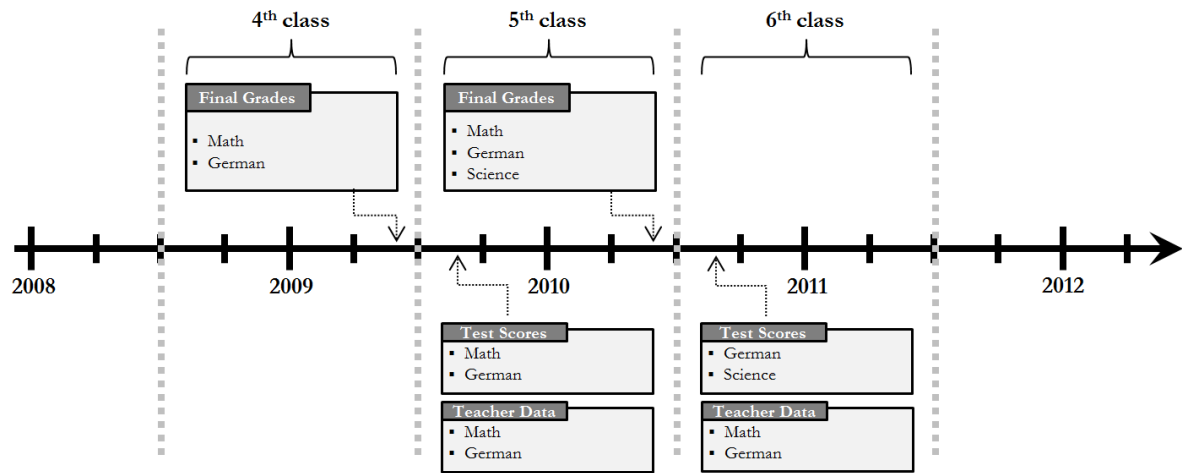
Table 4-1: Descriptive statistics on students' tests scores, grades, and background

	Female				Male			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Test scores and grades								
Math grade	-0.108	(0.981)	-3.8	1.5	0.106	(1.004)	-3.8	1.5
Math test score	-0.127	(0.994)	-4	3.5	0.128	(0.991)	-3	3.5
German grade	0.135	(0.959)	-4.1	1.7	-0.127	(1.018)	-4.1	1.7
German test score	-0.017	(0.904)	-3.9	2.7	0.025	(0.893)	-5.1	3.1
Science grade	0.020	(0.958)	-4.3	1.4	-0.016	(1.025)	-4.3	1.4
Scientific test score	-0.096	(0.964)	-2.6	6.7	0.096	(1.025)	-3.3	6.7
Student background								
DGCF (perceptual speed)	0.106	(0.981)	-3.2	3.6	-0.083	(1.007)	-3.3	3.6
DGCF (reasoning)	-0.025	(0.987)	-2.6	1.9	0.055	(1.006)	-2.6	1.9
Age	11.735	(0.779)	6.2	15	11.836	(0.807)	9.3	17
Born in Germany	0.964	(0.187)	0	1	0.955	(0.207)	0	1
Household size	4.424	(1.384)	0	15	4.463	(1.408)	0	15
PC availability	1.344	(0.527)	0	2	1.408	(0.531)	0	2
Life satisfaction	7.877	(2.301)	0	10	7.856	(2.336)	0	10
Freq: Sports	3.785	(1.051)	1	5	4.061	(1.085)	1	5
Freq: Reading	3.270	(1.345)	1	5	2.852	(1.405)	1	5
Number of books	3.859	(1.356)	1	6	3.824	(1.447)	1	6
Share of migrants in school	3.153	(1.290)	1	7	2.976	(1.326)	1	7
Friends' care for school	2.896	(1.147)	1	5	2.970	(1.220)	1	5
Non-cognitive skills								
SDQ-Scale: Prosocial behavior	7.867	(1.701)	0	10	6.857	(2.009)	0	10
SDQ-Scale: Problem behavior	2.275	(1.813)	0	10	2.548	(1.913)	0	10
Considerate	2.567	(0.516)	1	3	2.331	(0.564)	1	3
Likes to share things	2.600	(0.531)	1	3	2.393	(0.603)	1	3
Loner	1.557	(0.662)	1	3	1.615	(0.708)	1	3
Helpful	2.718	(0.485)	1	3	2.466	(0.589)	1	3
Has friends	2.845	(0.396)	1	3	2.816	(0.443)	1	3
Popular	2.262	(0.636)	1	3	2.305	(0.626)	1	3
Nice to younger children	2.661	(0.518)	1	3	2.457	(0.590)	1	3
Is teased	1.319	(0.580)	1	3	1.426	(0.645)	1	3
Helps others voluntarily	2.319	(0.565)	1	3	2.212	(0.601)	1	3
Gets along better with adults	1.508	(0.619)	1	3	1.642	(0.675)	1	3
Observations	4941				5216			

Note: All variables of the category 'Test scores and grades' are standardized to have a mean of zero and standard deviations of one with higher values indicating better performance. All variables of the category 'Student background' are when not stated otherwise self-explanatory. DGCF variables represent basic cognitive skills with higher values indicating better performance. PC availability is scaled with 0 (No PC), 1 (Shared PC) and 2 (Own PC). Life satisfaction is scaled from 0 (Very unhappy) to 10 (Very happy). Frequency of sports, frequency of reading, number of books, share of migrants and friends' care for school are scaled from 1 (Low, low frequency or low number) to 5, 6 or 7 (High, high frequency or high number). All variables in the 'Non-cognitive skills' section are reported from 1 (Not applicable) to 3 (Clearly applicable), except for the SDQ measures in prosocial and problem behavior that are reported from 0 (Clearly Applicable) to 10 (Not applicable).

Table 4-2: Estimated gender gaps in mathematics, German, and science

	(1)			(2)			(3)			(4)		
	Math	German	Science	Math	German	Science	Math	German	Science	Math	German	Science
Female	-0.185***	0.237***	0.0708	-0.193***	0.238***	0.0612	-0.159***	0.265***	0.0529	-0.197***	0.222***	0.0368
	(0.0271)	(0.0214)	(0.0445)	(0.0285)	(0.0218)	(0.0486)	(0.0381)	(0.0273)	(0.0681)	(0.0446)	(0.0289)	(0.0804)
Test score	0.334***	0.547***	0.222***	0.328***	0.528***	0.225***	0.284***	0.485***	0.191***	0.265***	0.429***	0.194***
	(0.0169)	(0.0212)	(0.0245)	(0.0173)	(0.0226)	(0.0270)	(0.0243)	(0.0282)	(0.0407)	(0.0259)	(0.0317)	(0.0397)
School FE		Yes			Yes			Yes			Yes	
Classroom FE		No			Yes			Yes			Yes	
Student info		No			No			Yes			Yes	
Non-cognitive		No			No			No			Yes	
Observations	4786	9313	2415	4786	9313	2415	2788	5790	1577	2342	4858	1322
R^2	0.419	0.298	0.277	0.461	0.367	0.364	0.534	0.430	0.475	0.571	0.481	0.536

Note: Dependent variable: standardized grade with a mean of zero and a standard deviation of one in the respective subject. The regressions were separately run with standard errors clustered by school and including the respective fixed effect indicated by 'School FE' or 'Classroom FE'. Standard errors are shown in parenthesis below the coefficients. 'Student info' includes two measures of intelligence as well as age, friends' care for schooling, migrant and socio-economic status, frequency of sports and reading, share of migrants in the classroom, computer availability at home and life satisfaction. 'Non-cognitive' includes results from two SDQ questionnaires and measures of being 'considerate', 'helpful', 'popular', 'nice to younger children' and 'teased by other children', as well as 'getting along better with adults than children', 'being a loner', 'having friends' and 'likes to share'. Stars denote significance of the estimates as follows: ***1%, **5%, and *10%.

5 The Impact of the Bologna Reform on Student Outcomes
   Evidence from Exogenous Variation in Regional Supply of Bachelor Programs in Germany[1]

5.1 Introduction

Higher education is generally perceived as becoming increasingly relevant in today's knowledge economies (Vandenbussche, Aghion, and Meghir 2006). In this regard, a country's future competitiveness relates to the productivity of its tertiary education system. The Bologna Reform was aimed at increasing the efficiency and attractiveness of higher education within European countries. In particular, policy-makers wanted to increase the mobility and employability of university students by introducing a homogeneous degree system based on two main cycles, the Bachelor/Master system (European Ministers of Education 2003). In Germany, this led to the abandoning of the hitherto single degree system. Since the Bachelor degree (the first cycle degree) can be obtained in less time than a traditional degree, the new degree system reduces the costs of earning a first tertiary education degree. This reduction in costs could be expected to increase enrollment and reduce dropout rates. Policy-makers also hoped that the harmonization of the degree structure across European countries would increase in particular international student mobility.

This chapter investigates to what extend the restructuring of the higher education degree system in Germany had the intended effects on students' mobility and employability. In particular, we analyze the effects of the reform on international

---

and national student mobility as one of the major policy goals. While any direct measures of labor market outcomes are not yet available, we also analyze the effects on outcomes which are potentially related to employability, such as dropout and internship participation: Dropping out of university may reduce an individual's employment opportunities, participating in internships may increase them. In addition, we investigate whether the reform had a negative impact on the study atmosphere as perceived by students to evaluate the concern of unintended side effects.

We exploit exogenous variation in the local availability of Bachelor programs to estimate causal effects of the reform on student outcomes in Germany. Due to the decentralized implementation of Bachelor degree programs in Germany, both old and new degree programs coexisted for several years leading, on the one hand, to the possibility to evaluate reform effects without confounding changes over time, but, on the other hand, to potentially endogenous sorting of students into old and new degree programs. To solve this endogeneity problem, we employ an instrumental variables approach by instrumenting enrollment into a Bachelor's program with the distance differential between an individual's nearest university with a Bachelor's and the nearest university with a traditional degree program.

We use a unique micro-level dataset on German high-school graduates of 2006 whom we observe in 2009. This dataset contains information on the place of high school of the individual which enables us to link these data to rich administrative data on university study programs in 2006 to employ our instrumental variables approach.

Our estimation results do not provide evidence that the reform had a significant effect on student mobility, dropout, and internship participation on average. However, we find a statistically significant negative effect on dropout for higher achieving students of about 10 percent and a borderline significant negative effect on dropout for females of about 9 percent. Furthermore, we find evidence that the reform had a positive impact on the study atmosphere as perceived by students.

We are not aware of any study that evaluates the effect of the Bologna Reform on student mobility, although this was one of the major policy goals. In a related study, Parey and Waldinger (2011) analyze the introduction of the ERASMUS program, which provides financial aid to students when going abroad, and find a significantly positive effect on international student mobility.

Existing research has mainly focused on the impact of the reform on enrollment and dropout rates with different findings across countries. Positive enrollment effects have been reported for Italy and Portugal (e.g. Cappellari, and Lucifora, 2009; Cardoso, 2008; Di Pietro, 2012), whereas no significant effect was found for Germany (Horstschräer, and Sprietsma, 2015). The evidence for dropout appears to be mixed even within a country.

The remainder of this part is structured as follows. Chapter 5.2 describes the Bologna Process and the changes it induced in the German higher education system in more detail. In Chapter 5.3, we discuss related literature. In Chapter 5.4, we describe the data and present our estimation strategy for the identification of causal effects. Chapter 5.5 contains our results. Chapter 5.6 concludes.

5.2 The Bologna Reform

On June 19, 1999 the Ministers of Education of 29 European countries met in the Italian city of Bologna to discuss a common strategy to promote the European higher education area. Set forth in the Bologna Declaration, the main objectives of the so-called Bologna Reform are to improve international competitiveness of the European higher education area, foster (international) mobility of students, teachers and researchers, and to strengthen the employability of the European university graduates. In particular, the latter goal gained much momentum in Germany triggered by a broad discussion about the efficiency of the German higher education system in the late 1990s and early 2000s.[2] Many scientists as well as politicians and employers criticized that the average German university student took too long to finish a degree, dropped out too frequently and was lacking important soft skills.

The universities of each member state were requested to introduce a system of easily readable and comparable degrees based upon two main cycles (see European Ministers of Education, 1999, p. 3) together with a unitary credit point system. In Germany, this led to the abandoning of the single-tier study programs and the respective degrees (called Diploma in some subjects and Magister in others) and the introduction of the two-tier Bachelor/Master system. Theoretically, the new two-tier

---

[2] For example, see Kultusministerkonferenz (1997) and Wissenschaftsrat (2000) for suggestions on how to improve the German higher education system.

system may offer some important advantages compared to the old single-tier system, but there may also be some disadvantages.

The Bachelor degree was thought of as a first academic degree which qualifies for direct labor market entry whereas the consecutive Master degree should provide a profound academic education for a scientific career. Since the Bachelor degree can be earned in less time compared to one of the traditional degrees, this should lower the costs of investing in tertiary education for individuals interested in acquiring basic academic skills and quickly entering the labor market. On the other hand, the Master degree, which requires the successful completion of a Bachelor's degree, offers a more specialized education, but students typically have to commit themselves to an overall longer duration of study than before.[3]

A two-tier system also makes it possible to offer Master programs which do not require a Bachelor's degree in the same subject which increases the options for students within the new system and, therefore, its attractiveness.[4] However, it is not clear to what extend Bachelor and Master degrees qualify for distinct employment positions. In practice, both Bachelor and Master graduates might compete for the same job offer. This may reduce the value of the Bachelor degree, since Bachelor graduates obtained less human capital than Master graduates. In fact, there is evidence that more than 72 percent of the students choose to obtain a Master's degree upon successful completion of the Bachelor's degree (Heine 2012).

The adoption process varied substantially across European countries: England, for instance, already had a two-tier Bachelor/Master system in place and had to carry out only minor adjustments. In Italy, the new system was introduced simultaneously at all universities in 2001. Portugal opted for a decentralized introduction of the new degrees and required its universities to switch to the new system at some point between 2006 and 2008. In Germany, universities were free to choose any point in time between 2000 and 2010 to introduce the new degree system. It was agreed upon that the introduction process should be completed by 2010. In Germany, this goal

---

3 The usual duration of Bachelor programs is three years, that of Master programs two years. Traditional programs took four to five years.

4 In fact, an explicit goal of the Bologna Reform also was the promotion of interdisciplinary study programs (European Ministers of Education 2003).

was widely achieved, with a few exceptions.[5] In 2003, less than 5 percent of all departments had adopted the new degree whereas by 2008 almost 90 percent had completely switched to offering Bachelor degrees (see Horstschräer and Sprietsma, 2015, p. 1).

The Bologna agreement did not provide any distinct implementation rules with regard to contents of the new degree programs. This led to a fairly heterogeneous adoption. Some departments tried to set up new programs that were specifically tailored to the shorter study period of the Bachelor cycle. Others continued to offer the same program and only replaced the old with the new degree which ultimately led to a tighter schedule of teaching (Winter et al. 2010).

5.3 Related Literature

The existing evidence on the effects of the Bologna Reform on student outcomes is rather scarce, although it induced large changes in the tertiary education systems of many European countries. This circumstance is most likely due to a lack of adequate data sources and compelling strategies to identify causal effects. Cappellari and Lucifora (2009), for instance, estimate the effect of the Bachelor introduction in Italy on enrollment and dropout rates using a simple before-after comparison, thereby ignoring any potential biases from time trends as well as confounding factors that may have occurred together with the implementation of the Bologna Reform and that may have had an effect on the enrollment decision. Di Pietro (2012) re-evaluates their analysis by employing a difference-in-differences approach. The author argues that the Bologna Reform was primarily targeted towards individuals from less advantaged social backgrounds, so that this subgroup constitutes the treatment group. He identifies individuals as belonging to the treatment group when neither of their parents have a university degree. Individuals with at least one parent with a university degree constitute the control group.[6] In order to capture the effect from time trends in enrollment, the author uses four cohorts of high school leavers, two before and two after the Bachelor introduction in Italy in 2001.

---

5 For example, neither of the medicine departments introduced the new degrees. Likewise, law departments were still o ering traditional degree programs by 2010.

6 Although not explicitly stated in the paper, the author most likely refers to the fact that in theory the Bachelor introduction reduced the cost of investing in higher education, because it takes less time to earn a rst degree so that the investment becomes pro table for individuals at the margin of investing.

While this approach is more refined than a simple before-after comparison, it hinges on the assumption that the Bachelor introduction did not affect individuals from the control group. In fact, it is plausible to assume that the Bologna Reform also influenced individuals from the control group in their decision to enroll in higher education as it introduced a considerable amount of flexibility as described in Chapter 2.2. If this also motivated more individuals from the control group to enroll in higher education, the reform effect is underestimated. Cappellari and Lucifora (2009) conclude that the reform increased enrollment by 15 percent, whereas Di Pietro (2012) estimates a reform effect of 7 percent.

Two further studies attempt to gauge the effect on dropout rates in the Italian context based on mainly descriptive evidence: D'Hombres (2007) finds significant lower dropout rates among post-reform cohorts of university students, whereas Boero et al. (2005) find no evidence of reduced dropout. Finally, Bratti et al. (2006) analyze the extent to which the reform had an impact on study programs. They analyze data from a single Italian university department and conclude that it became easier for students to pass first-year courses. Cardoso (2008) and Portela et al. (2009) analyze students' demand for study programs in Portugal. They find that departments which introduced the Bachelor degree were more often chosen by first-year students than those which remained offering a traditional degree program.

In a recent study, Horstschräer and Sprietsma (2013) analyze the effect of the Bologna Reform on enrollment and dropout rates in Germany. They employ axed effects panel model to analyze administrative data on the department level from 1998 to 2008. Overall, they do not find any effect of the Bachelor introduction on neither enrollment nor dropout rates. However, results appear to differ by subjects. In English Language, German Language as well as Computer Sciences the Bachelor introduction seems to have had a positive enrollment effect, whereas in Mechanical Engineering and Electrical Engineering the effect is negative. Due to the decentralized introduction of Bachelor programs in Germany, i.e. old and new degree programs coexisted for several years, this result is likely to reflect students' selection into one or the other degree program. For the analysis of dropout rates a similar picture emerges. For Biology, the estimated effect is positive, whereas it is negative for Business Administration, English Language Studies, and German Language Studies. Unfortunately, the authors are not able to distinguish between students who quit studying and those who change subject or university.

Mühlenweg (2010) tries to answer the question whether studying in a Bachelor's program affected students' satisfaction. Controlling for observable student

characteristics, she concludes that the satisfaction of students in Bachelor programs is slightly higher compared to their peers in traditional degree programs.

Finally, there are a few studies in the spirit of our IV approach exploiting proximity to a specific (treatment) location as a source of exogenous variation. Originating in labor economics, studies in that field exploit distance measures orthogonal to unobserved individual heterogeneity to investigate for example labor market returns of further training programs (Mallar 1979), years of schooling (Card 1995) and type of college and its degree's completion (Maluccio 1998). In other areas, studies have used the distance to the nearest nursery (Attanasio et al. 2013) or hospital (Baiocchi et al. 2010; McClellan al. 1994) to evaluate their causal impact.

5.4   Data and descriptive statistics

For our analysis, we use a cross-section from a rich panel dataset on German high track leavers who graduated in 2006. We observe the individuals in December of 2009, i.e. three and a half years after graduating from high school.[7] The survey is conducted by the DZHW and offers some important advantages for analyzing the effects of the Bologna Reform on student outcomes. First, the dataset allows us to analyze several outcome variables related to the policy goals of the Bologna Reform. Second, it contains information on a student's place of high school (zip code) which enables us to merge information on German universities and their degree programs in 2006. This information is needed for our instrumental variable approach which is described in detail in the following chapter.

The dataset contains information on a student's international and national mobility, i.e. whether he/she went abroad for interim studies and whether he/she changed his/her university. It also contains information on whether a student dropped out or not and whether he/she did an internship while enrolled. The last two variables are likely to play a role for an individual's employability. Dropping out of tertiary education may signal a lower ability so that this outcome should be negatively correlated with labor market success. On the other hand, internship participation may increase an individual's chances on the labor market. Since most students were

---

7 The individuals were originally sampled in 2005, when they were still in school. However, all our outcome variables are contained in the 2009 questionnaire.

still enrolled at the time of the interview, we cannot observe any direct labor market outcomes yet.

To relate the student information to the tertiary education supply in 2006, we obtained an administrative dataset containing information on the universe of German higher education institutions and their degree programs in 2006 from the German Rectors' Conference (Hochschulrektorenkonferenz (HRK)).[8] For every institution of higher education, the dataset includes information on its type and degree programs (including the subject and the degree) offered in the winter term of 2006/07.[9] Based on the awarded degree, we constructed a categorical variable on the university-subject level: 1 if only a Bachelor's program was offered, 2 if only a traditional degree program was offered, 3 if both a Bachelor and a traditional degree program were offered, and 0 if the subject was not offered at all.[10] Since correspondence courses are not bound to a specific location, we did not consider them in our analysis.

Based on the university's address, we geocoded all universities and used QGIS to calculate the air-line distance between an individual's place of high school and the universities. We merged the university data to our student dataset using the zip code of the high school location. This resulted in a student-university-level dataset, where each student was matched with 409 university observations. In addition, we obtained information on the GDP, population, and size of each county in Germany in 2006 from the regional statistics database of the Federal Statistical Once. From this, we calculated the population density as inhabitants per km^2 and the GDP per capita for each county and merged these variables at the high school county level to our individual data. This enables us to control for regional characteristics of a

---

8 Some universities (especially universities of applied sciences) have departments that are located in different regions/towns, which is not accounted for in the original data. As our identification strategy is based on regional variation in the availability of degree programs, it was important to ensure that the location of the departments was exact. Therefore, in some cases, we had to manually check and add information on the exact location of a department.

9 There are three basic types of higher education institutions in Germany. One is rather research oriented, called university, the other is rather applied, called university of applied sciences, and the third offers only art subjects, called art college . The funding of these institutions can either be public, private, or clerical.

10 The traditional degree category comprises all old degree types such as Diplom, Magister, and Staatsexamen. We also included teaching degrees if it was clear to which category (old or new) they belonged.

student's origin (place of high school). We consider students enrolled in the 20 most popular subjects as of the winter term 2006/07 (Statistisches Bundesamt, 2007, p. 46) which accounts for 68.3 percent of all students in the original dataset after dropping medical students.[11]

Our final dataset has a sample size of 1626 students, who enrolled in either a Bachelor's or a traditional degree program between the winter term 2006/07 and the winter term 2007/08. We have information on a student's demographic characteristics such as a student's gender, age, nationality, and father's and mother's education. Information is also provided on a student's grade point average in the high school exit exam and the type of the high school leaving certificate (general or subject specific university entrance diploma). We observe the subject in which a student enrolled and the semester of enrollment. For 1471 students we also observe the first university attended.

Table 1 provides summary statistics of our data. 56 percent of the students in our sample are enrolled in a Bachelor's degree program. The other 44 percent are enrolled in a traditional degree program. Students are 23 years old on average and have a high school GPA of 2.9 on a scale from 1 (lowest) to 4 (highest). 35 percent of the students enrolled in a subject within the area of social sciences, 26 percent within natural sciences, 21 percent within engineering, and 18 percent within language and culture studies. The nearest university is on average 23 km away from the high school location of the student. This distance varies considerably within a range from 0 to 115 km.[12] By the time we observe the individuals in 2009, 7.3 percent had gone abroad for interim studies, 2.3 percent had changed their university, 3.1 percent had dropped out, and 20.2 percent had done an internship. On average, a student's assessment of the study atmosphere is 3.9 on a scale from 1 (lowest) to 5 (highest).

---

11 The information on the degree programs was raw data, meaning that it indicated the specific title of the program. We were very cautious in categorizing them into subjects so as to avoid coding errors. As this was a time-intensive process, we focused our analysis on the 20 most popular subjects which are: business administration, law, German, medicine, mechanical engineering, computer sciences, economics, industrial engineering and management, electrical engineering, mathematics, biology, English, educational science, architecture, psychology, chemistry, physics, construction engineering, business informatics, political science. Since there were no Bachelor programs in medicine, we omitted this subject.

12 Due to data protection rules, we had to aggregate our distance measure in intervals of 5 km starting with zero.

For many variables there is a significant difference between students enrolled in a Bachelor's versus a traditional degree program. It is likely that a considerable fraction of these differences is due to student selection into old and new degree programs. The large differences in the fields of study also reflect variation in the timing of the introduction of the new degree system across departments. On average, programs in social sciences were changed earlier to the new degree system compared to programs in language and culture studies.

Most of our outcomes increase in probability with the time since enrollment. For example, students who enrolled earlier than others are more likely to have gone abroad by the time we observe the students in our data. In our sample, 68.5 percent of the students enrolled in the winter term 2006/07, 4.1 percent enrolled in the summer term 2007, and 27.4 percent enrolled in the winter term 2007/08. The later enrollment rates are mainly caused by male students due to the military/civilian service requirement at that time. 46 percent of the male students in our sample began their studies in the fall of 2007. To capture time effects from differential enrollment dates, we control for time of enrollment in all of our regressions.

5.5   Estimation strategy

To investigate the relationship between studying in a Bachelor's degree program and student outcomes of individual i from federal state m in subject l, we consider a model of the following form:

$$Y_{ilm} = \alpha_1 + \beta_1 Bachelor_i + X_i'\gamma_1 + \delta_l + \mu_m + \epsilon_{ilm} \tag{5-1}$$

Y denotes our respective outcome of interest: going abroad, change of university, dropout, internship, and students' satisfaction with the study atmosphere. Change of university includes only changes within a subject and degree program. This means that students who changed universities because they wanted to study a different subject or degree type are excluded. *Bachelor* indicates studying in a Bachelor's degree program compared to in a traditional degree program and X is a vector of covariates that includes student demographic characteristics, information about parents education, and information about the location of the high school. We include subject dummies (δ) in order to account for unobserved heterogeneities between subjects. We also include state dummies with respect to the high school of an individual (μ). These are necessary because schooling policies, such as high school curricula, are set at the state level and can have a substantial impact on graduates' preparation for tertiary education. To account for interdependence of

observations within a university, we cluster the standard errors on the attended university level.

The parameter of interest in the equation above is β_1 which is supposed to capture the effect of studying in a Bachelor's degree program on the respective student outcome. Estimating the equation by OLS, however, may yield biased estimates. Although controlling for potentially confounding influences can reduce the threat of biases, one can easily think of unobserved heterogeneities that can have influenced the selection of students into new or old degree programs. For example, since the new degree programs were intended to facilitate the transferability of course credits, it is possible that students with a higher taste for mobility choose to enroll in Bachelor's programs. In a regression with going abroad as our outcome variable, β_1 would be biased upwards, since the unobserved variable 'taste for mobility' is positively correlated with studying in a Bachelor's program.

To solve the problem of omitted variable bias we apply an instrumental variables (IV) approach that exploits regional variation in the supply of Bachelor and traditional degree programs. Due to the decentralized introduction of the Bachelor degree system in Germany under which university departments were free to choose when to implement the Bachelor, both degree systems coexisted for many years. Our IV approach is based on the idea that most students choose to attend a local university so that it is the local education supply which matters to them. Figure 1 shows the distribution of the distance between a student's place of high school and his/her first attended university in our sample. The graph reveals that, indeed, most students decide to enroll at a university close to their place of origin.[13]

We construct our instrument as the difference in distances between the nearest public university with a Bachelor's program and the nearest public university with a traditional program in a student's subject. We condition our instrument on a student's subject for two reasons: First, in 2006, almost all universities had introduced the Bachelor degree in at least one subject. Constructing the instrument on the university rather than the subject (department) level would result in almost no variation in the instrumental variable which is needed to identify a causal effect.

---

13 This fact is also established in a number of other studies: For example, Spieß, and Wrohlich (2010) investigate the relationship between the distance to the nearest university from a student's home and university attendance in Germany and find a negative correlation.

Second, there is evidence that the personal interest in a particular subject is by far the most important determinant of the decision where and what to study (Heine et al. 2005, 2008).[14] We further restrict our university data to public institutions since 95 percent of all students in the winter term 2006/07 enrolled in a public institution (Statistisches Bundesamt, 2007, p. 60). However, we do provide a robustness check using all universities (including private and clerical institutions) in the distance calculation.

Let *MinDist_trad$_i$* denote the air-line distance between student i's place of high school and the nearest university with a traditional degree program in student *i*'s subject.[15] Accordingly, let *MinDist_ba$_i$* denote the air-line distance between student *i*'s place of high school and the nearest university with a Bachelor's degree program in student *i*'s subject. The difference of these two distance measures yields our instrumental variable:

$$IV \equiv \textit{Distance differential}_i = \textit{MinDist\_trad}_i - \textit{MinDist\_ba}_i. \qquad (5\text{-}2)$$

The distance differential can be thought of as a measure of the regional supply with a Bachelor's program relative to a traditional program.[16] Thus, our first stage is given by the following equation:

$$Bachelor_{ilm} = \alpha + \beta Distance differential_i + X_i'\gamma + \delta_l + \mu_m + \varepsilon_{lm} \qquad (5\text{-}3)$$

The intuition is as follows: The nearer the university with a Bachelor's degree program relative to the university with a traditional degree program in student *i*'s subject, the likelier it is that student *i* enrolled in a Bachelor's degree program. Spiess and Wrohlich (2010) have already shown that the distance from home to a university plays a significant role in the educational choice of high school graduates

---

14 Hachmeister (2007, p. 58) provide suggestive evidence that almost 95 percent of German students choose their subject before their university location.

15 We use the place of high school to calculate our distance measure, because we do not have exact information on a student's place of residence at the time he/she finishes school. In practice, this should not make a big difference since most students attend a school close to their home.

16 A relative distance measure is also used in an instrumental variables approach in Oosterbeek et al. (2010) to estimate the effect of entrepreneurship education on entrepreneurship skills and motivation.

in Germany, hence anything but a likewise effect on choice of type of program would be surprising.

Our IV approach identifies a local average treatment effect (Angrist, and Imbens, 1994), i.e. the effect of the Bachelor introduction for individuals for whom distance matters. These individuals have higher transaction costs of moving to a faraway university than on average and thus prefer to attend an institution which is close to their home. In an attempt to reveal some basic traits of potential compliers in our sample, we divide the students into quartiles according to the distance between the place of high school and the first attended university. As can be observed from Table 2, students who stay rather close to their hometown (Column 1) have on average worse high school GPA scores compared to more mobile students and also are from lower educated families.

We also estimate the effects of the Bologna Reform using a modified version of the instrument described above. Because the German higher education system comprises two main types of higher education institutions (i.e. universities which are rather research oriented and universities which are rather applied), it might be the case that many students only consider studying at one specific type of university. Since our data provides information on a student's first attended university, we are able to calculate the distance differential based on the type of the university attended. Students who only consider studying at one type of university may constitute a different complier group, so that we do not expect the results to remain unchanged. Figure 2 shows density plots of our two instruments. There is substantial variation in both instruments, although for most students the nearest universities that offer new and old degree programs in the chosen subject are located rather close to each other. The last two rows in Table 1 contain summary statistics of our instruments. The average distance differential for IV1 is -1.24 km, for IV2 -7.3 km. IV2 denotes the instrument in which we account for the type of university attended. Students who enrolled in a traditional degree program have a negative distance differential on average which means that the nearest Bachelor university is farther away than the nearest university with a traditional degree program. For students who enrolled in a Bachelor's degree program the distance differential is positive on average which means that the Bachelor university is closer.

5.6   Results

Our headline results are presented in Table 3. All regressions are based on linear probability models with the exception of the categorical outcome variable 'satisfaction' which ranges from 1 (lowest) to 5 (highest). The standard errors in all estimated models are clustered on the attended university level. The first stage F-statistics in all IV specifications are sufficiently large to reject weak instrument concerns. We further divide the student population into subgroups to investigate heterogeneous effects on different subpopulations. In particular, we analyze heterogeneities by gender and high school GPA. Reduced-form estimation results are contained in Table A.1.

Table 5 displays the results of OLS regressions for the respective outcome. Column 1 shows the effect of the Bologna Reform on international student mobility. Participation in a Bachelor's degree program has a small, positive, but insignificant effect of 0.02. Other explanatory variables have the expected signs. For example, better students, as measured by the high school GPA, have a higher probability of going abroad. A higher socio-economic background, as measured by the educational attainment of the parents, also increases the probability of going abroad. Time of enrollment is negatively correlated with going abroad reflecting the time effect of later enrollment.

Results for the effect on national student mobility (change of university) are reported in Column 2. Participation in a Bachelor's degree program has no effect on the probability of changing universities. Germans have a 3 percent higher probability of changing universities compared to immigrants.

OLS estimates further suggest that participating in a Bachelor's degree program has no effect on dropout (Column 3) or internship participation (Column 4). Better students have a significantly lower dropout probability (2.7 percent per 1 point better high school GPA) and a higher, although insignificant, probability of doing an internship. Later enrollment significantly lowers the probability of having done an internship by the time the students are observed. A one year later enrollment is associated with an 11 percent lower probability of having done an internship.

Column 5 shows the effect of participating in a Bachelor's degree program on a student's satisfaction with the study atmosphere. Results suggest that students in a Bachelor's program are more content than students in a traditional degree program, although the effect is rather small. On a scale from 1 to 5, the effect is 0.11. Female

students are on average less content than male students and younger students are on average more content than older students.

Table 4 provides first stage regression results for IV1 and IV2. The potentially endogenous variable *Bachelor* is regressed on the instrument and further explanatory variables. Each specification in columns 1 to 6 (IV1) includes additional covariates and fixed effects. Column 6 and 7 report estimates of IV1 and IV2, respectively, in our preferred specification. The F-statistic for IV1 is 18.86 and for IV2 22.42. Throughout all specifications, the estimated effect of the instrument on participating in a Bachelor's degree program is highly significant and fairly robust. The probability increases by 1.3 to 2.9 percent with every 10 km depending on the respective specification. This confirms our hypothesis that the nearer a department with a Bachelor's degree program relative to a department with a traditional degree program the more likely it is that a student enrolled in a Bachelor's program. We find a highly significant effect of 0.0029 in a univariate regression of the Bachelor indicator variable on IV1 (Column 1). The inclusion of student controls, region controls, and state of high school fixed effects does not change the effect. Only the inclusion of subject fixed effects reduced the estimate to 0.0017 for IV1 and 0.0013 for IV2.

Results also show that the type of high school degree plays a crucial role whether a student enrolled in a Bachelor's or a traditional degree program. Students who obtained a subject specific or vocational university entrance diploma (i.e. study options are either limited to certain subjects or to the type of university) have a higher probability to enroll in a Bachelor's program compared to students with a general university entrance diploma. It may be that these students are attracted to the Bachelor degree due to the shorter duration of study. Results also show that the time of enrollment is a major determinant of enrolling in a Bachelor's degree program. Since the availability of Bachelor's programs increased over time whereas the availability of traditional programs decreased, the probability to enroll in a Bachelor's program increased by 26 to 29 percent for one year later enrollment.

As discussed above, OLS results are potentially biased by omitted variables. Table 3 presents our IV results using IV1 and IV2 in separate regressions for all outcomes. As mentioned above, we do not expect identical results from both IVs due to potentially different complier groups. Using IV1, we estimate a local average treatment effect (LATE) for students for whom the local tertiary education supply matters. Using IV2, we estimate a LATE for students who, in addition, make a more conscious decision about the type of university they want to enroll at. This group of

students is likely to be better informed about their expected study conditions compared to the complier group of IV1.

Columns 1 and 2 contain our estimates of the effect of the Bologna Reform on international mobility. Results show no effect when using IV1 as an instrument for enrolling in a Bachelor's degree program. However, using IV2, we find a positive effect of 0.17 which is almost statistically significant at the 10 percent level. Since we most probably estimate different LATEs with IV1 and IV2, it may be the case that students who make a deliberate choice regarding the type of university are also more able to take advantage of the new homogeneous degree system which was intended to facilitated the transfer of course credits between universities. The estimates for high school GPA and parent education background have the expected sign in both IV regressions. A one point better high school GPA leads to a 4 to 5 percent higher probability of going abroad. Better educated parents also increase the probability of going abroad, although the effect is small.

IV point estimates for the impact of the reform on national mobility (change of university) indicate that there may be a small positive effect of roughly 2 percent in both IV specifications (Columns 3 and 4). However, standard errors increased substantially compared to the OLS estimations so that the effect is not statistically significant. Since IV is less efficient than OLS, the increase in the size of the standard errors is a common phenomenon in IV approaches. In addition, it is worth mentioning that our sample size is rather low with less than 1500 observations and about 200 cluster. It may be that the results show the true effect, however, we cannot make a definite statement. Intuitively, it makes sense that the new degree system may have increased the probability of changing universities because of the easier transferability of course credits.

The effect on dropout is shown in Column 5 and 6. Compared to the OLS result which indicates no effect of the reform on dropout, IV results suggest that the dropout probability decreased by 1.5 to 3.8 percent. Again, standard errors are large for the reasons discussed above so that the effect is statistically insignificant. High school GPA has a negative impact on dropout which is in line with the common view that better students are more likely to finish their studies.

Columns 7 and 8 show our IV estimates of the effect of the reform on the probability of doing an internship. Whereas the OLS estimate is zero, the IV estimates are 0.04 and 0.07. Both estimates are not statistically significant due to large standard errors. Unfortunately, we do not have enough information to what extend the introduction

of the new degree system caused changes in study conditions that might have facilitated doing an internship.

Columns 9 and 10 contain the results for the effect of the reform on students' satisfaction with the study atmosphere. The estimate is 0.35 in the IV1-regression and 1.25 in the IV2-regression. The latter is statistically significant at the 10 percent level. Both estimates are larger than the OLS estimate of 0.11. This suggests that the Bologna Reform had, in fact, a positive impact on the study atmosphere as perceived by students. The larger point estimate in our IV2-regression might again reflect the specific effect for students who deliberately chose one type of university.[17]

As our IV estimates do not provide clear evidence due to a lack of statistical significance, we cannot definitively state that the Bologna Reform had an impact on student mobility, dropout, and internship participation. However, IV point estimates slightly deviate from OLS point estimates. OLS estimates might be biased due to omitted variables, whereas IV estimates are unbiased but imprecisely estimated.

We also estimated the effects of the Bologna Reform on the outcomes using an unconditional distance differential as the instrumental variable. In particular, we included private and clerical institutions in the distance calculations. In comparison, IV1 is calculated using only public universities. Due to the fact that only 5 percent of all students enroll at private and clerical universities, the relevance of the unconditional instrument is lower compared to IV1. The first stage F-statistic is approximately 16 for this instrument, compared to 19 for IV1 and 22 for IV2. Nevertheless, we find very similar results to our IV1 specification.

It might be that certain subgroups of our student population were affected differently by the introduction of the new degree system. To explore the impact of the Bologna Reform on student outcomes in more detail, we estimate separate effects by gender and high school GPA. We do not find pronounced effect heterogeneities for our considered outcomes except for dropout (Table 6). For female students, we find that the reform reduced the dropout probability by about 9

17 The differences in the estimates are not due to differences in sample size. Due to missing information in the variable indicating the university attended, IV2-regressions are based on a lower sample size than IV1-regressions. However, restricting the IV1-regressions to the sample used in the IV2-regressions yields almost identical results for IV1.

percent. When IV1 is used as the instrument, the effect is almost statistically significant at the 10 percent level. In comparison to the IV results, OLS yields an estimated effect of zero as in the full sample. The instruments are highly relevant for females with first stage F-statistics of almost 24. For males, the F-statistics are insufficiently large so that we cannot make a statement for this subgroup.

We also find differential effects for students with a high school GPA above versus below the median of 2.9. For high achievers (GPA > 2.9), we find that the reform significantly (IV2) reduced the dropout probability by 9 to 10 percent. For low achievers (GPA < 2.9), point estimates are positive but not statistically significant.

The identifying assumption of our estimation strategy is that the distance differential is uncorrelated with any observable or unobservable covariates which are not included in the regression. This requires that the Bachelor introduction was geographically random conditional on covariates included in the regressions. As stated earlier, the introduction of the Bachelor degree system occurred on rather heterogeneous grounds, because there was no common introduction plan. There is evidence that the variation in pace of introduction within a subject area was mainly caused by external, political pressure and not due to university or department specific factors like quality, finance or prestige (Krücken et al., 2005). However, individuals from rural areas are likely to have larger distance differentials than individuals from urban areas due to the lower density of universities in rural areas. To account for this possibility, we control for regional characteristics of a student's place of high school which we believe to capture potentially spurious correlation between our instrument and geographic differences.

In Table 7, we provide suggestive evidence on the exogeneity of our instruments. The table shows results from regressions of the instruments on student characteristics and our regional controls. We do not find significant correlations between a student characteristic and the instruments, except for a weakly significant relationship between IV2 and the gender variable. Most notably, there is no correlation between a student's high school GPA and our instruments. Column 9 contains estimates from a regression of the instruments on all student characteristics. Their joint significance can be rejected as indicated by the p-values.

5.7   Conclusion

This study investigates the impact of the Bologna Reform on student mobility, dropout, internship participation, and a student's satisfaction with the study

atmosphere in Germany using survey data from 2009 on German high track leavers who graduated in 2006. To account for the potentially endogenous sorting of individuals into new and old degree programs at the time of enrollment, we use an instrumental variables approach based on the nearest universities that offer a Bachelor's and a traditional degree program in a student's subject. In particular, we use the distance differential between the nearest university with a Bachelor's and the nearest university with a traditional degree program in a student's subject as an instrument for participation in a Bachelor's degree program.

Overall, we do not find a significant effect from studying in a Bachelor's degree program on student mobility, dropout, and internship participation. However, we find a significantly negative effect on dropout for higher achieving students of about 10 percent and an almost significantly negative effect on dropout for females of about 9 percent. Results further indicate that the reform had a positive effect on a student's satisfaction with the study atmosphere.

It is important to emphasize that our results should be interpreted as short-term effects. Since we analyze students that were among the first cohorts to enroll in a Bachelor's program, our estimates are likely to reflect also the circumstances of the introduction of the new degree system. In many cases the new degree structure was applied to existing programs without much adjustments in study content. As the new study programs are gradually being improved and adjusted to the new two-tier degree structure, effects may differ for more recent cohorts. One should also keep in mind, that our IV approach identifies a local average treatment effect for individuals for whom distance matters. This means that the results are not easily transferable to more mobile students.

Future research should explore the mediating channels of the reform in more detail and try to disentangle the effects of the new, homogeneous, two-tier degree structure from effects related to adjustments in study content. To fully assess the implications of the reform, especially in light of further policy advice, it is crucial to also evaluate the reform effects on direct labor market outcomes, such as wages or unemployment probability. Once appropriate data become available, one could use the IV strategy presented in this study to estimate causal effects of the reform on these outcomes.

Figure 5-1: Distribution of distance to university attended
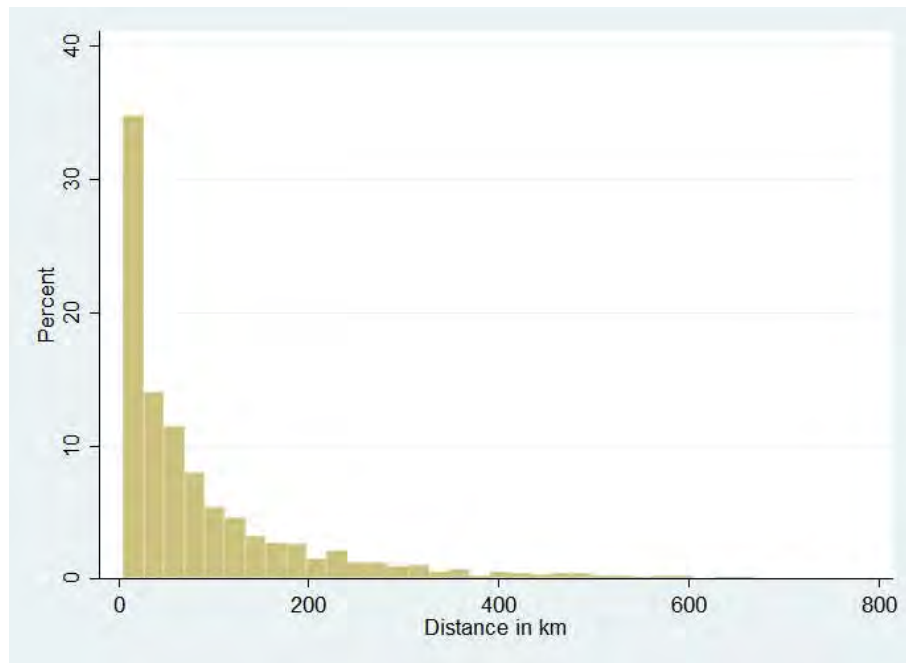
Note: The figure shows the distribution of distances between a student's place of high school and the first university attended in our data.

Figure 5-2: Density distribution of distance differential
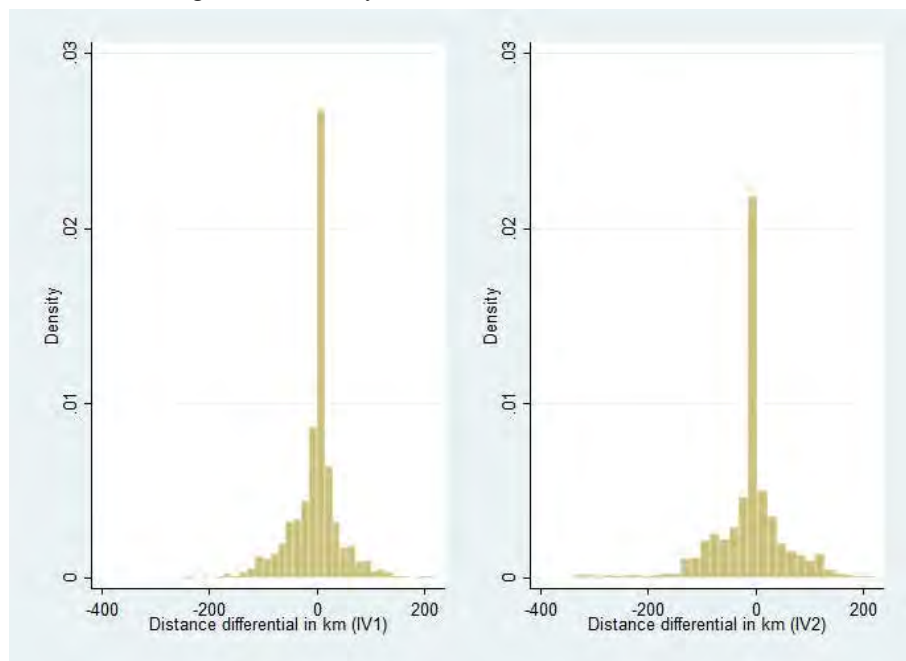
Note: The figure shows the density distributions of our instruments. IV1 represents the distance differential between the nearest public university with a traditional degree program and the nearest public university with a Bachelor's program in a student's subject. IV2 represents the distance differential between the nearest public university with a traditional degree program and the nearest public university with a Bachelor's program in a student's subject while additionally accounting for the type of university a student enrolled at.

Table 5-1: Summary statistics

	Full Sample			Traditional			Bachelor			Difference [SE]
	Mean (SD)	Min/Max	Obs	Mean (SD)	Min/Max	Obs	Mean (SD)	Min/Max	Obs	
Bachelor	0.561	0	1626							
	(0.496)	1								
Student Characteristics										
Female	0.601	0	1626	0.667	0	714	0.549	0	912	0.117***
	(0.490)	1		(0.472)	1		(0.498)	1		[0.024]
Year of birth	1986.274	1968	1624	1986.283	1968	714	1986.267	1969	910	0.016
	(1.305)	1989		(1.311)	1989		(1.302)	1988		[0.065]
German	0.974	0	1622	0.965	0	712	0.981	0	910	-0.016**
	(0.159)	1		(0.184)	1		(0.135)	1		[0.008]
High school GPA	2.917	1	1621	2.934	1	712	2.904	1	909	0.030
	(0.586)	4		(0.590)	4		(0.583)	4		[0.029]
Type of HS degree	0.087	0	1626	0.057	0	714	0.110	0	912	-0.052***
	(0.282)	1		(0.233)	1		(0.313)	1		[0.014]
Father's education	3.592	0	1626	3.634	0	714	3.559	0	912	0.075
	(1.396)	5		(1.360)	5		(1.422)	5		[0.069]
Mother's education	3.490	0	1626	3.517	0	714	3.469	0	912	0.048
	(1.280)	5		(1.259)	5		(1.296)	5		[0.064]
Enrollment WS 2006	0.685	0	1623	0.791	0	714	0.602	0	909	0.190***
	(0.465)	1		(0.407)	1		(0.490)	1		[0.022]
Enrollment SS 2007	0.041	0	1623	0.055	0	714	0.031	0	909	0.024**
	(0.199)	1		(0.227)	1		(0.173)	1		[0.010]
Enrollment WS 2007	0.274	0	1623	0.154	0	714	0.367	0	909	-0.213***
	(0.446)	1		(0.361)	1		(0.482)	1		[0.021]
Distance to next univ. in km	23.444	5	1626	23.018	5	714	23.777	5	912	-0.759
	(17.999)	115		(17.473)	115		(18.404)	105		[0.894]
Regional Characteristics										
GDP/C in Euro	29684.307	13542	1445	30770.451	13956	623	28861.110	13542	822	1909.341***
	(10285.564)	79039		(11534.295)	79039		(9149.057)	79039		[561.586]
Population density	818.197	39	1626	849.775	43	714	793.476	39	912	56.299
	(915.877)	4171		(925.772)	4171		(907.798)	4171		[45.869]

Continued on next page

Continued from previous page

	Full Sample			Traditional			Bachelor			
	Mean (SD)	Min/Max	Obs	Mean (SD)	Min/Max	Obs	Mean (SD)	Min/Max	Obs	Difference [SE]
Area of Study										
Language and Culture	0.179 (0.383)	0 1	1626	0.248 (0.432)	0 1	714	0.125 (0.331)	0 1	912	0.122*** [0.0195]
Social Sciences	0.348 (0.477)	0 1	1626	0.280 (0.449)	0 1	714	0.401 (0.490)	0 1	912	-0.121*** [0.0233]
Natural Sciences	0.261 (0.440)	0 1	1626	0.286 (0.452)	0 1	714	0.242 (0.429)	0 1	912	0.0433** [0.0220]
Engineering	0.212 (0.409)	0 1	1626	0.186 (0.390)	0 1	714	0.231 (0.422)	0 1	912	-0.0450** [0.0201]
Outcomes										
Going Abroad	0.073 (0.260)	0 1	1626	0.076 (0.265)	0 1	714	0.070 (0.256)	0 1	912	0.005 [0.013]
Change of university	0.023 (0.151)	0 1	1626	0.031 (0.173)	0 1	714	0.018 (0.131)	0 1	912	0.013* [0.008]
Dropout	0.031 (0.174)	0 1	1536	0.033 (0.178)	0 1	675	0.030 (0.171)	0 1	861	0.002 [0.009]
Internship	0.202 (0.402)	0 1	1626	0.210 (0.408)	0 1	714	0.196 (0.397)	0 1	912	0.014 [0.020]
Satisfaction	3.881 (0.899)	1 5	1612	3.769 (0.913)	1 5	711	3.969 (0.878)	1 5	901	-0.200*** [0.045]
Instruments										
IV1	-1.236 (44.703)	-240 220	1626	-14.545 (42.389)	-240 155	714	9.183 (43.713)	-170 220	912	-23.728*** [2.147]
IV2	-7.301 (64.595)	-335 220	1471	-29.688 (65.581)	-335 155	705	13.303 (56.328)	-185 220	766	-42.991*** [3.200]

Notes: The left part shows descriptive statistics on the whole student sample, the middle part on students enrolled in a traditional degree and the right part on students enrolled in a Bachelor's degree program. The last column shows the difference in means between both programs with standard errors [SE] in brackets. Standard deviations (SD) are shown in parentheses. To the right of each mean are the minima (Min) and maxima (Max) of each variable, as well as the number of observations (Obs). High school GPA ranges from 1 (worst) to 4 (best). Type of HS (high school) degree indicates subject specific *Fachabitur* versus general university entrance diploma *Abitur*. Father's and mother's edu- cation are categorical variables ranging from 0 (No degree) to 5 (University degree). Distance to next university is measured in kilometers and indicates the distance to the next university within a student's field of study. GDP/C indicates the gross domestic product per capita in Euro within a county. Population density is measured as inhabitants per km^2. All outcome variables are binary, with the exception of satisfaction which ranges from 1 (unhappy) to 5 (happy). Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5-2: Descriptive statistics by distance to university attended

	1st quart. 0 - 20 km Mean (SD)	2nd quart. 20 - 50 km Mean (SD)	3rd quart. 50 - 120 km Mean (SD)	4th quart. 120 - 670 km Mean (SD)
Female	0.627	0.608	0.637	0.567
	(0.484)	(0.489)	(0.482)	(0.496)
Year of birth	1986.193	1986.222	1986.417	1986.289
	(1.602)	(1.488)	(0.871)	(1.174)
German	0.972	0.963	0.980	0.975
	(0.164)	(0.189)	(0.139)	(0.157)
High school GPA	2.829	2.855	2.901	3.030
	(0.594)	(0.574)	(0.598)	(0.563)
Type of HS degree	0.087	0.105	0.082	0.070
	(0.282)	(0.307)	(0.274)	(0.256)
Father's education	3.471	3.480	3.682	3.857
	(1.437)	(1.442)	(1.350)	(1.304)
Mother's education	3.398	3.392	3.470	3.803
	(1.317)	(1.267)	(1.283)	(1.204)
Enrollment WS 2006	0.725	0.715	0.699	0.612
	(0.447)	(0.452)	(0.460)	(0.488)
Enrollment SS 2007	0.048	0.037	0.042	0.039
	(0.214)	(0.189)	(0.201)	(0.195)
Enrollment WS 2007	0.227	0.248	0.259	0.348
	(0.419)	(0.432)	(0.439)	(0.477)
Distance to next univ. in km	12.437	25.298	31.451	27.444
	(9.812)	(13.014)	(21.311)	(20.971)
Observations	437	352	355	356

Notes: The table contains descriptive statistics by the distance between a student's place of high school and the first university attended. Each column shows means and standard deviations of student characteristics within quartiles of the distance distribution.

Table 5-3: IV results for the effect of the Bologna Reform on student outcomes

Dep. Var.:	(1) Going Abroad		(2) Change of University		(3) Dropout		(4) Internship		(5) Satisfaction	
	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2
Bachelor	-0.0102	0.1713	0.0266	0.0172	-0.0148	-0.0377	0.0409	0.0755	0.3451	1.2487**
	(0.0914)	(0.1157)	(0.0560)	(0.0644)	(0.0609)	(0.0751)	(0.1592)	(0.1597)	(0.3633)	(0.3541)
Female	0.0014	-0.0016	0.0047	0.0019	0.0120	0.0146	0.0373	0.0354	-0.1035*	-0.1230
	(0.0190)	(0.0222)	(0.0111)	(0.0124)	(0.0118)	(0.0137)	(0.0257)	(0.0268)	(0.0572)	(0.0750)
Year of birth	0.0063*	0.0055	0.0012	0.0015	-0.0117	-0.0117	0.0129	0.0169**	0.0448**	0.0471**
	(0.0033)	(0.0037)	(0.0032)	(0.0035)	(0.0085)	(0.0089)	(0.0080)	(0.0075)	(0.0206)	(0.0219)
German	0.0266	0.0263	0.0293**	0.0316***	-0.0170	-0.0146	0.0304	0.0303	0.1156	0.1026
	(0.0373)	(0.0380)	(0.0082)	(0.0087)	(0.0414)	(0.0431)	(0.0503)	(0.0530)	(0.1658)	(0.1805)
High school GPA	0.0420**	0.0495**	-0.0131	-0.0142*	-0.0259**	-0.0273**	0.0246	0.0424**	0.0217	-0.0004
	(0.0115)	(0.0122)	(0.0084)	(0.0085)	(0.0130)	(0.0135)	(0.0205)	(0.0213)	(0.0492)	(0.0549)
Type of HS degree	0.0078	-0.0160	0.0019	0.0070	0.0509	0.0560	-0.0765*	-0.1033**	-0.0533	-0.1713
	(0.0210)	(0.0264)	(0.0174)	(0.0204)	(0.0335)	(0.0391)	(0.0455)	(0.0517)	(0.1009)	(0.1266)
Father's education	0.0060	0.0051	0.0028	0.0037	-0.0038	-0.0043	0.0102	0.0065	0.0020	0.0072
	(0.0045)	(0.0048)	(0.0034)	(0.0037)	(0.0042)	(0.0045)	(0.0085)	(0.0086)	(0.0184)	(0.0226)
Mother's education	0.0105*	0.0111*	0.0007	0.0017	0.0037	0.0044	-0.0001	-0.0026	0.0224	0.0338
	(0.0057)	(0.0059)	(0.0034)	(0.0034)	(0.0041)	(0.0045)	(0.0086)	(0.0089)	(0.0225)	(0.0265)
Enrollment SS 2007	-0.0261	-0.0361	0.0109	0.0108	-0.0353***	-0.0357***	-0.0579	-0.0776	-0.0484	-0.0772
	(0.0317)	(0.0342)	(0.0256)	(0.0281)	(0.0099)	(0.0113)	(0.0474)	(0.0504)	(0.1436)	(0.1602)
Enrollment WS 2007	-0.0135	-0.0678*	-0.0057	-0.0068	0.0072	0.0153	-0.1249**	-0.1434***	-0.0372	-0.3223**
	(0.0278)	(0.0376)	(0.0191)	(0.0221)	(0.0194)	(0.0248)	(0.0467)	(0.0529)	(0.1049)	(0.1252)
Distance to next university	-0.0003	-0.0006	-0.0002	-0.0003	0.0003	0.0004	0.0000	-0.0001	-0.0002	-0.0021
	(0.0005)	(0.0006)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0009)	(0.0009)	(0.0016)	(0.0019)
Region controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State of high school FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1434	1292	1434	1292	1365	1227	1434	1292	1424	1282
Cluster	231	197	231	197	228	194	231	197	229	195
1st stage F-Stat	18.8608	22.4182	18.8608	22.4182	16.5960	21.8354	18.8608	22.4182	18.6729	21.6952

Notes: Dependent variable as indicated in the first row. 1 to 4 are binary outcomes, 5 is categorical ranging from 1 (lowest) to 5 (highest). Standard errors are clustered on the attended university level. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Dep. Var.: Bachelor (1=yes; 0=no)

Table 5-4: First stage results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IV1	0.0029***	0.0029**	0.0029**	0.0029**	0.0028***	0.0017**	
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	
IV2							0.0013**
							(0.0003)
Female		-0.0169	-0.0178	-0.0323	-0.0293	0.0265	0.0373
		(0.0362)	(0.0360)	(0.0364)	(0.0365)	(0.0316)	(0.0340)
Year of birth		0.0060	0.0055	0.0066	0.0105	0.0023	0.0018
		(0.0114)	(0.0114)	(0.0110)	(0.0110)	(0.0099)	(0.0108)
German		0.0909	0.0902	0.0810	0.0607	0.0355	0.0367
		(0.0746)	(0.0747)	(0.0775)	(0.0752)	(0.0632)	(0.0650)
High school GPA		0.0078	0.0071	0.0246	0.0338	0.0412**	0.0276
		(0.0250)	(0.0251)	(0.0238)	(0.0239)	(0.0200)	(0.0201)
Type of HS degree		0.1551**	0.1548**	0.1419**	0.1639***	0.0851	0.1165**
		(0.0461)	(0.0458)	(0.0466)	(0.0475)	(0.0530)	(0.0548)
Father's education		-0.0111	-0.0106	-0.0072	-0.0066	-0.0035	-0.0005
		(0.0088)	(0.0088)	(0.0095)	(0.0094)	(0.0085)	(0.0088)
Mother's education		-0.0014	-0.0009	-0.0104	-0.0152	-0.0068	-0.0060
		(0.0108)	(0.0108)	(0.0113)	(0.0114)	(0.0096)	(0.0104)
Enrollment SS 2007		-0.0547	-0.0532	-0.0127	-0.0137	0.0058	0.0313
		(0.0712)	(0.0709)	(0.0728)	(0.0738)	(0.0552)	(0.0592)
Enrollment WS 2007		0.2535**	0.2527**	0.2699**	0.2660***	0.2578**	0.2869**
		(0.0326)	(0.0325)	(0.0338)	(0.0343)	(0.0324)	(0.0338)
Distance to next university in km			0.0007	-0.0002	-0.0006	0.0019**	0.0016**
			(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Region controls	No	No	No	Yes	Yes	Yes	Yes
State of high school FE	No	No	No	No	Yes	Yes	Yes
Subject FE	No	No	No	No	No	Yes	Yes
Observations	1625	1613	1613	1434	1434	1434	1292
Cluster	241	241	241	231	231	231	197
R^2	0.0691	0.1349	0.1355	0.1485	0.1705	0.3243	0.3259
F-Statistic	87.8161	88.1004	89.2101	75.6017	55.6276	18.8608	22.4182

Notes: Dependant variable indicates studying in a Bachelor's degree program (1) versus a traditional degree program (0). All specifications employ a linear probability model (LPM). Standard errors are clustered on the attended university level. All explanatory variables including the indicated regional controls are described in Table 1. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5-5: OLS results for the effect of the Bologna Reform on student outcomes

Dep. Var.:	(1) Going Abroad	(2) Change of Univer-	(3) Dropout	(4) Internship	(5) Satisfaction
Bachelor	0.0213	-0.0008	0.0011	0.0001	0.1050**
	(0.0160)	(0.0095	(0.0104)	(0.0264)	(0.0491)
Female	0.0007	0.0053	0.0115	0.0383	-0.0985*
	(0.0194)	(0.0114	(0.0117)	(0.0255)	(0.0557)
Year of birth	0.0062*	0.0013	-0.0116	0.0130	0.0454**
	(0.0033)	(0.0032	(0.0087)	(0.0082)	(0.0216)
German	0.0251	0.0306	-0.0179	0.0323	0.1266
	(0.0375)	(0.0079	(0.0422)	(0.0498)	(0.1681)
High school GPA	0.0407***	-0.0120	-0.0266**	0.0263	0.0319
	(0.0111)	(0.0080	(0.0126)	(0.0199)	(0.0501)
Type of HS degree	0.0048	0.0045	0.0494	-0.0726	-0.0296
	(0.0200)	(0.0178	(0.0347)	(0.0445)	(0.0903)
Father's education	0.0062	0.0027	-0.0037	0.0100	0.0009
	(0.0045)	(0.0035	(0.0042)	(0.0085)	(0.0182)
Mother's education	0.0106*	0.0006	0.0037	-0.0003	0.0213
	(0.0056)	(0.0033	(0.0042)	(0.0087)	(0.0223)
Enrollment SS 2007	-0.0262	0.0110	-0.0356***	-0.0577	-0.0448
	(0.0320)	(0.0263	(0.0104)	(0.0490)	(0.1457)
Enrollment WS 2007	-0.0214	0.0012	0.0030	-0.1146***	0.0236
	(0.0157)	(0.0117	(0.0119)	(0.0232)	(0.0613)
Distance to next university	-0.0004	-0.0001	0.0003	0.0001	0.0003
	(0.0005)	(0.0003	(0.0004)	(0.0008)	(0.0015)
Region controls	Yes	)	Yes	Yes	Yes
State of high school FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes
Observations	1434	1434	1365	1434	1424
Cluster	231	231	228	231	229
R^2	0.0278	0.0507		0.0893	0.0733

Notes: Dependent variable as indicated in the first row. 1 to 4 are binary outcomes, 5 is categorical ranging from 1 (lowest) to 5 (highest). Standard errors are clustered on the attended university level. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5-6: Heterogeneous effects by gender and high school GPA

Dep. Var.: Dropout

	OLS	IV1	IV2	OLS	IV1	IV2
		Females			Males	
Bachelor	0.0084	-0.0891	-0.0891	-0.0115	0.3473	0.1588
	(0.0123)	(0.0592)	(0.0902)	(0.0167)	(0.3151)	(0.1777)
Observations	818	818	752	547	547	475
F-Statistic		23.8050	23.5201		1.0927	3.5440
		Above median high school GPA			Below median high school GPA	
Bachelor	0.0054	-0.0865	-0.1004*	0.0003	0.1159	0.0154
	(0.0128)	(0.0635)	(0.0598)	(0.0180)	(0.1354)	(0.1002)
Observations	763	763	667	602	602	560
F-Statistic		12.1722	11.4985		7.2376	20.5845
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Region controls	Yes	Yes	Yes	Yes	Yes	Yes
State of high school FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Binary dependent variable for dropout (1=yes, 0=no). The upper panel shows estimation re- sults of studying in a Bachelor's degree program for females and males, respectively. The lower panel shows estimation results of studying in a Bachelor's degree program for students with a high school GPA above and below the median of 2.9. Standard errors are clustered on the attended university level. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5-7: Correlations between the instruments and observed student characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Var.: IV1										
Female	-3.0892									-3.0298
	(2.6251)									(2.8143)
Year of birth		-0.4467								-0.1298
		(0.9100)								(0.9188)
German			6.2182							7.0965
			(5.5634)							(5.5223)
High school GPA				-3.1721						-3.4519
				(2.2791)						(2.2261)
Type of HS degree					4.2643					2.7147
					(5.0773)					(5.2131)
Father's education						-0.4812				-0.9399
						(0.9387)				(0.9758)
Mother's education							0.7401			1.5774
							(1.0029)			(0.9977)
Enrollment SS 2007								-4.9525		-5.3893
								(5.0699)		(5.1991)
Enrollment WS 2007									-0.0001	-1.1676
									(0.0262)	(2.8270)
Regional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1434	1434	1434	1434	1434	1434	1434	1434	1434	1434
p-value	0.2405	0.6240	0.2649	0.1653	0.4019	0.6087	0.4613	0.3297	0.9981	0.3253
Dep. Var.: IV2										
Female	-7.5960*									-5.5751
	(4.2644)									(4.5275)
Year of birth		-0.8896								-0.1935
		(1.3827)								(1.2619)
German			6.0936							6.6868
			(10.0159)							(9.5799)
High school GPA				-2.6498						-2.1950
				(3.4218)						(3.2497)
Type of HS degree					14.8800					12.8274
					(13.6675)					(13.9312)
Father's education						-0.4358				-1.2285
						(1.3750)				(1.4786)
Mother's education							1.2284			2.4206*
							(1.3066)			(1.3332)
Enrollment SS 2007								-5.2798		-4.1309
								(8.2141)		(8.3264)
Enrollment WS 2007									0.0328	1.3906
									(0.0294)	(4.1515)
Regional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1292	1292	1292	1292	1292	1292	1292	1292	1292	1292
p-value	0.0764	0.5207	0.5436	0.4396	0.2776	0.7516	0.3483	0.5211	0.2645	0.5303

Notes: The table contains results from regressions of the instrumental variables (IV1 and IV2) on observed student characteristics. In Columns 1 to 9, the p-value indicates the level of significance of the explanatory variable. In Column 10, the p-value indicates the level of joint significance of all explanatory variables. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

5.8 Appendix

Table 5-8: Reduced forms

Dep. Var.:	(1) Going Abroad		(2) Change of University		(3) Dropout		(4) Internship		(5) Satisfaction	
IV1	-0.0000 (0.0002)		0.0000 (0.0001)		-0.0000 (0.0001)		0.0001 (0.0003)		0.0006 (0.0006)	
IV2		0.0002 (0.0001)		0.0000 (0.0001)		-0.0001 (0.0001)		0.0001 (0.0002)		0.0017*** (0.0004)
Female	0.0011 (0.0193)	0.0048 (0.0213)	0.0054 (0.0114)	0.0025 (0.0127)	0.0115 (0.0118)	0.0125 (0.0132)	0.0384 (0.0254)	0.0382 (0.0269)	-0.0952* (0.0556)	-0.0801 (0.0581)
Year of birth	0.0063* (0.0034)	0.0058 (0.0036)	0.0013 (0.0032)	0.0015 (0.0035)	-0.0116 (0.0086)	-0.0116 (0.0092)	0.0130 (0.0082)	0.0170** (0.0077)	0.0456** (0.0218)	0.0494** (0.0234)
German	0.0263 (0.0377)	0.0326 (0.0393)	0.0302** (0.0080)	0.0322** (0.0083)	-0.0177 (0.0421)	-0.0162 (0.0434)	0.0318 (0.0500)	0.0331 (0.0528)	0.1274 (0.1690)	0.1455 (0.1723)
High school GPA	0.0416*** (0.0112)	0.0542*** (0.0116)	-0.0120 (0.0079)	-0.0137 (0.0086)	-0.0265** (0.0127)	-0.0285** (0.0136)	0.0263 (0.0199)	0.0445** (0.0211)	0.0361 (0.0498)	0.0338 (0.0530)
Type of HS degree	0.0069 (0.0201)	0.0040 (0.0227)	0.0041 (0.0173)	0.0090 (0.0194)	0.0496 (0.0349)	0.0513 (0.0388)	-0.0730* (0.0441)	-0.0945* (0.0489)	-0.0226 (0.0910)	-0.0198 (0.0925)
Father's education	0.0061 (0.0045)	0.0050 (0.0048)	0.0028 (0.0035)	0.0037 (0.0038)	-0.0037 (0.0042)	-0.0042 (0.0046)	0.0100 (0.0085)	0.0064 (0.0088)	0.0007 (0.0181)	0.0070 (0.0193)
Mother's education	0.0106* (0.0056)	0.0101* (0.0060)	0.0006 (0.0033)	0.0016 (0.0034)	0.0038 (0.0042)	0.0046 (0.0047)	-0.0004 (0.0088)	-0.0030 (0.0091)	0.0200 (0.0221)	0.0262 (0.0233)
Enrollment SS 2007	-0.0261 (0.0322)	-0.0307 (0.0351)	0.0110 (0.0263)	0.0113 (0.0288)	-0.0356*** (0.0104)	-0.0376*** (0.0117)	-0.0576 (0.0489)	-0.0753 (0.0523)	-0.0439 (0.1459)	-0.0281 (0.1473)
Enrollment WS 2007	-0.0161 (0.0155)	-0.0187 (0.0167)	0.0011 (0.0109)	-0.0019 (0.0117)	0.0032 (0.0119)	0.0039 (0.0133)	-0.1144*** (0.0215)	-0.1217*** (0.0228)	0.0518 (0.0583)	0.0366 (0.0621)
Distance to next univ. in km	-0.0003 (0.0005)	-0.0003 (0.0005)	-0.0001 (0.0003)	-0.0003 (0.0002)	0.0003 (0.0004)	0.0003 (0.0004)	0.0001 (0.0008)	-0.0000 (0.0008)	0.0005 (0.0015)	-0.0001 (0.0016)
Region controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State of high school FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1434	1292	1434	1292	1365	1227	1434	1292	1424	1282
Cluster	231	197	231	197	228	194	231	197	229	195
R^2	0.0868	0.0279	0.0291	0.0507	0.0528	0.0894	0.1058		0.0716	0.0819

Notes: Dependent variable as indicated in the first row. 1 to 4 are binary outcomes, 5 is categorical ranging from 1 (lowest) to 5 (highest). Standard errors are clustered on the attended university level. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

6 References

Aaronson, D., L. Barrow, And W. Sander. 2007. "Teachers And Student Achievement In The Chicago Public High Schools." *Journal Of Labor Economics* 25(1): 95–135.

Abadie, Alberto, And Javier Gardeazabal. 2003. "The Economic Costs Of Conflict: A Case Study Of The Basque Country." *The American Economic Review* 93(1): 113–32.

Abdulkadiroguglu, Atila, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, And Parag A Pathak. 2011. "Accountability And Flexibility In Public Schools: Evidence From Boston's Charters And Pilots." *The Quarterly Journal Of Economics* 126(2): 699–748.

Altonji, Joseph G., And Charles R. Pierret. 2001. "Employer Learning And Statistical Discrimination." *The Quarterly Journal Of Economics* 116(1): 313–50.

Anders, Yvonne, Mareike Kunter, Martin Brunner, Stefan Krauss, And Jürgen Baumert. 2010. "Diagnostische Fähigkeiten Von Mathematiklehrkräften Und Ihre Auswirkungen Auf Die Leistungen Ihrer Schülerinnen Und Schüler." *Psychologie In Erziehung Und Unterricht* (3): 175–93.

Anderson, Theodore Wilbur, And Cheng Hsiao. 1982. "Formulation And Estimation Of Dynamic Models Using Panel Data." *Journal Of Econometrics* 18(1): 47–82.

Angrist, Joshua D. "The Perils Of Peer Effects." *Labour Economics* 30 (2014): 98-108.

Angrist, Joshua, Eric Bettinger, And Michael Kremer. 2006. "Long-Term Educational Consequences Of Secondary School Vouchers: Evidence From Administrative Records In Colombia." *The American Economic Review* 96(3): 847–62.

Angrist, Joshua D, And Guido W Imbens. 1994. "Identification And Estimation Of Local Average Treatment Effects." *Econometrica* 62(2): 467–75.

Angrist, Joshua D, And Victor Lavy. 1999. "Using Maimonides' Rule To Estimate The Effect Of Class Size On Student Achievement." *Quarterly Journal Of Economics* 114(2): 533–75.

Angrist, Joshua D, Parag A Pathak, And Christopher R Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5(4): 1–27.

Angrist, Joshua D, And Jörn-Steffen Pischke. 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, New Jersey.

Ashenfelter, Orley. 1978. "Estimating The Effect Of Training Programs On Earnings." The *Review Of Economics And Statistics*: 47–57.

Aslam, M., And G. Kingdon. 2011. "What Can Teachers Do To Raise Pupil Achievement?" *Economics Of Education Review* 30(3): 559–74.

Attanasio, Orazio P, Vincenzo Di Maro, And Marcos Vera-Hernández. 2013. "Community Nurseries And The Nutritional Status Of Poor Children. Evidence From Colombia." *The Economic Journal* 123(571): 1025–58.

Baiocchi, Mike, Dylan S Small, Scott Lorch, And Paul R Rosenbaum. 2010. "Building A Stronger Instrument In An Observational Study Of Perinatal Care For Premature Infants." *Journal Of The American Statistical Association* 105(492): 1285–96.

Baker, Bruce D. 2001. "Can Flexible Non-Linear Modeling Tell Us Anything New About Educational Productivity?" *Economics Of Education Review* 20(1): 81–92.

Baumert, J., And M. Kunter. 2011. "Das Mathematikspezifische Wissen Von Lehrkraeften, Kognitive Aktivierung Im Unterricht Und Lernfortschritte Von Schuelerinnen Und Schuelern." In Professionelle Kompetenz Von Lehrkraeften: Ergebnisse Des Forschungsprogramms COACTIV, Eds. M Kunter, J Baumert, W Blum, Klusmann U, S Krauss, And M Neubrand. Muenster: Waxmann, 163–92.

Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, S. Krauss, M. Neubrandt, And Y.-M. Tsai. 2010. "Teachers' Mathematical Knowledge, Cognitive Activation In The Classroom, And Student Progress." *American Educational Research Journal* 47(1): 133–80.

Becker, Gary S. 1962. "Investment In Human Capital: A Theoretical Analysis." *The Journal Of Political Economy*: 9–49.

Becker, Gary S. 1964. Handbook Of The Economics Of Education. Reflections On The Economics Of Education. Elsevier B.V.

Betts, Julian, And John E Roemer. 2005. "Equalizing Opportunity For Racial And Socioeconomic Groups In The United States Through Educational Finance Reform." Department Of Economics, UCSD.

Bishop, John H. 1997. "The Effect Of National Standards And Curriculum-Based Exams On Achievement." *The American Economic Review* 87(2): 260–64.

Björklund, Anders, And Kjell G Salvanes. 2011. "Chapter 3-Education And Family Background: Mechanisms And Policies. Volume 3 Of Handbook Of The Economics Of Education, 201--247."

Blossfeld, Hans-Peter, Hans-Günther Roßbach, And Jutta Von Maurice. 2011. *Zeitschrift Für Erziehungswissenschaft - Sonderheft Education As A Lifelong Process: The German National Educational Panel Study (NEPS).*

BMBF. 2012. Systematisierung Der Lehrerforschung Und Verbesserung Ihrer Datenbasis. Berlin.

Boero, Gianna, Tiziana Laureti, And Robin Naylor. 2005. "An Econometric Analysis Of Student Withdrawal And Progression In Post-Reform Italian Universities." *Crenos Working Paper 2005/04.*

Böhlmark, Anders, And Mikael Lindahl. 2015. "Independent Schools And Long-Run Educational Outcomes: Evidence From Sweden's Large-Scale Voucher Reform." *Economica* 82(327): 508–51.

Bowles, Samuel. 1970. "Towards An Educational Production Function." *Education, Income, And Human Capital, NBER,* 11–70.

Bowles, Samuel, And Herbert Gintis. 1975. "The Problem With Human Capital Theory--A Marxian Critique." *The American Economic Review* 65(2): 74–82.

Boyd, D., H. Lankford, S. Loeb, J. Rockoff, And J. Wyckoff. 2008. "The Narrowing Gap In New York City Teacher Qualifications And Its Implications For Student Achievement In High-Poverty Schools." *Journal Of Policy Analysis And Management* 27(4): 793–818.

Bratti, Massimiliano, Chiara Broccolini, And Stefano Staffolani. 2006. "Is `3+2' Equal To 4? University Reform And Student Academic Performance In Italy." Quaderni Di Ricerca 251. Ancona: Universita Politecnica Delle Marche.

Brunello, Giorgio, Daniele Fabbri, And Margherita Fort. 2013. "The Causal Effect Of Education On Body Mass : Evidence From Europe." *Journal Of Labor Economics* 31(1): 195–223.

Burgess, Simon, And Ellen Greaves. 2013. "Test Scores, Subjective Assessment, And Stereotyping Of Ethnic Minorities." *Journal Of Labor Economics* 31(3): 535–76.

Cantrell, S., J. Fullerton, T. J. Kane, And D. O. Staiger. 2008. National Board Certification And Teacher Effectiveness: Evidence From A Random Assignment Experiment. Cambridge, MA.

Cappellari, Lorenzo, And Claudio Lucifora. 2009. "The 'Bologna Process' And College Enrollment Decisions." *Labour Economics* 16(6): 638–47.

Card, David. 1990. "The Impact Of The Mariel Boatlift On The Miami Labor Market." *Industrial & Labor Relations Review* 43(2): 245–57.

Card, David. 1995. "Using Geographic Variation In College Proximity To Estimate The Return To Schooling." In *Aspects Of Labour Market Behaviour: Essays In Honour Of John Vanderkamp*, Eds. Loizos Nicolaou Christofides, E Kenneth Grant, And Robert Swidinsky. Toronto: University Of Toronto Press, 201–22.

Card, David. 1999. "The Causal Effect Of Education On Earnings." *Handbook Of Labor Economics* 3: 1801–63.

Card, David, And Alan B Krueger. 1994. "Minimum Wages And Employment: A Case Study Of The Fast-Food Industry In New Jersey And Pennsylvania." *The American Economic Review* 84(4): 772–93.

Cardoso, Ana Rute. 2008. "Demand For Higher Education Programs: The Impact Of The Bologna Process." *Cesifo Economic Studies* 54(2): 229–47.

Carneiro, Pedro, Claire Crawford, And Alissa Goodman. 2007. "The Impact Of Early Cognitive And Non-Cognitive Skills On Later Outcomes."

Carneiro, Pedro, And James J Heckman. 2002. "The Evidence On Credit Constraints In Post-Secondary Schooling." *The Economic Journal* 112(482): 705–34.

Chamberlain, Gary. 1982. "Multivariate Regression Models For Panel Data." *Journal Of Econometrics* 18.

Chetty, Raj, John N. Friedman, And Jonah E. Rockoff. "Measuring The Impacts Of Teachers I: Evaluating Bias In Teacher Value-Added Estimates." *The American Economic Review* 104.9 (2014): 2593-2632.

Chetty, Raj, John N. Friedman, And Jonah E. Rockoff. "Measuring The Impacts Of Teachers II: Teacher Value-Added And Student Outcomes In Adulthood." *The American Economic Review* 104.9 (2014): 2633-2679.

Chingos, Matthew M., And Paul E. Peterson. 2011. "It's Easier To Pick A Good Teacher Than To Train One: Familiar And New Results On The Correlates Of Teacher Effectiveness." *Economics Of Education Review* 30(3): 449–65.

Clotfelter, C. T., H. F. Ladd, And J. L. Vigdor. 2006. "Teacher-Student Matching And The Assessment Of Teacher Effectiveness." *Journal Of Human Resources,* 41(4): 778–820.

Coleman, James Samuel. 1966. 2 *Equality Of Educational Opportunity [Summary Report*. US Department Of Health, Education, And Welfare, Office Of Education.

Cornwell, Christopher, David B. Mustard, And Jessica Van Parys. 2013. "Noncognitive Skills And The Gender Disparities In Test Scores And Teacher Assessments: Evidence From Primary School." *Journal Of Human Resources* 48(1): 236–64.

D'Hombres, B. 2007. "The Impact Of The Bologna Process On University Dropouts: Evidence From Italy." *JRC Scientific And Technical Reports 22733, The European Commission, Ispra, Italy*.

Dee, T. S. 2005. "A Teacher Like Me: Does Race, Ethnicity, Or Gender Matter?" *American Economic Review* 95(2): 158–65.

Dee, T. S. 2007. "Teachers And The Gender Gaps In Student Achievement." *Journal Of Human Resources* (September 2006).

Dehejia, Rajeev H, And Sadek Wahba. 1999. "Causal Effects In Nonexperimental Studies: Reevaluating The Evaluation Of Training Programs." *Journal Of The American Statistical Association* 94(448): 1053–62.

Dignath, C., And G. Buettner. 2008. "Components Of Fostering Self-Regulated Learning Among Students: A Meta-Analysis On Intervention Studies At Primary And Secondary School Level." *Metacognition And Learning* 3: 231–64.

Domar, Evsey D. 1946. "Capital Expansion, Rate Of Growth, And Employment." *Econometrica, Journal Of The Econometric Society*: 137–47.

Epstein, Roy J. 2014. 165 *A History Of Econometrics*. Elsevier.

European Ministers Of Education. 1999. "The Bologna Declaration Of 19 June 1999. Joint Declaration Of The European Ministers Of Education." *Bologna: European Ministers Of Education*.

European Ministers Of Education. 2003. "Realizing The European Higher Education Area." *Berlin: European Ministers Of Education*.

Fischer, Mira, And Bernhard Enzi. 2016. *Classroom Composition And Academic Self-Concept In Educational Production*, Mimeo.

Gardner, Howard. 1974. "The Arts And Human Development."

Goldhaber, D. 2002. "The Mystery Of Good Teaching." *Education Next* 2(1): 50–55.

Goldhaber, D., And E. Anthony. 2007. "Can Teacher Quality Be Effectively Assessed? National Board Certification As A Signal Of Effective Teaching." *Review Of Economics And Statistics* 89(1): 134–50.

Goldhaber, Dan, And Duncan Dunbar Chaplin. 2015. "Assessing The 'Rothstein Falsification Test': Does It Really Show Teacher Value-Added Models Are Biased?" *Journal Of Research On Educational Effectiveness* 8(1): 8–34.

Goldhaber, Dan D, And Dominic J Brewer. 1997. "Why Don't Schools And Teachers Seem To Matter? Assessing The Impact Of Unobservables On Educational Productivity." *Journal Of Human Resources*: 505–23.

Goldin, Claudia, Lawrence F. Katz, And Ilyana Kuziemko. 2006. "The Homecoming Of American College Women: The Reversal Of The College Gender Gap." *NBER Working Paper Series* 12139.

Grossbard-Shechtman, Shoshana, And Christopher K Clague. 2002. *The Expansion Of Economics: Toward A More Inclusive Social Science*. ME Sharpe.

Guarino, Cassandra M., Mark D. Reckase, And Jeffrey M. Wooldridge. "Can Value-Added Measures Of Teacher Performance Be Trusted?." *Education Finance And Policy* (2014).

Hachmeister, Cort-Denis, Maria E Harde, And Markus F Langer. 2007. "Einflussfaktoren Der Studienentscheidung. Eine Empirische Studie Von CHE Und EINSTIEG." *Arbeitspapier Nr. 95*.

Hanna, Rema, And Leigh Linden. 2009. "Measuring Discrimination In Education." *NBER Working Paper Series* 15057.

Hanushek, Eric A. "The Economic Value Of Higher Teacher Quality." *Economics Of Education Review* 30.3 (2011): 466-479.

Hanushek, E. A., And S. G. Rivkin. 2006. "Teacher Quality." In *Handbook Of The Economics Of Education*, Eds. E A Hanushek And F Welch. Amsterdam: North Holland, 1051–78.

Hanushek, Eric. 1971. "Teacher Characteristics And Gains In Student Achievement: Estimation Using Micro Data." *The American Economic Review* 61(2): 280–88.

Hanushek, Eric A. 1979. "Conceptual And Empirical Issues In The Estimation Of Educational Production Functions." *Journal Of Human Resources*: 351–88.

Hanushek, Eric A. 2003. "The Failure Of Input-Based Schooling Policies." *The Economic Journal* 113(485): F64–F98.

Hanushek, Eric A, Stephen J Machin, And Ludger Woessmann. 2016. *Handbook Of The Economics Of Education*. Elsevier.

Hanushek, Eric A, And Steven G Rivkin. 2010. "Generalizations About Using Value-Added Measures Of Teacher Quality." *American Economic Review* 100(2): 267–71.

Hanushek, Eric A, And Ludger Woessmann. 2008. "The Role Of Cognitive Skills In Economic Development." *Journal Of Economic Literature*: 607–68.

Hanushek, Eric A, And Ludger Wößmann. 2006. "Does Early Tracking Affect Educational Inequality And Performance? Differences-In-Differences Evidence Across Countries." *Economic Journal* 116(510).

Harris, D. N., And T. R. Sass. 2011. "Teacher Training, Teacher Quality And Student Achievement." *Journal Of Public Economics* 95(7-8): 798–812.

Harrod, Roy F. 1939. "An Essay In Dynamic Theory." *The Economic Journal* 49(193): 14–33.

Hattie, John. 2013. *Visible Learning: A Synthesis Of Over 800 Meta-Analyses Relating To Achievement*. Routledge.

Heckman, James J, And Yona Rubinstein. 2001. "The Importance Of Noncognitive Skills: Lessons From The GED Testing Program." *The American Economic Review* 91(2): 145–49.

Heine, Christoph. 2012. "Übergang Vom Bachelor- Zum Masterstudium." *HIS: Forum Hochschule, 7*.

Heine, Christoph, Heike Spangenberg, Jochen Schreiber, And Dieter Sommer. 2005. "Studienanfänger 2003/04 Und 2004/05 Bildungswege, Motive Der Studienentscheidung Und Gründe Der Hochschulwahl." *HIS: Kurzinformation, A 15*.

Heine, Christoph, Julia Willich, Heidrun Schneider, And Dieter Sommer. 2008. "Studienanfänger Im Wintersemester 2007/08. Wege Zum Studium, Studien- Und Hochschulwahl, Situation Bei Studienbeginn." *HIS: Forum Hochschule, 16*.

Heyneman, Stephen P, And William A Loxley. 1983. "The Effect Of Primary-School Quality On Academic Achievement Across Twenty-Nine High-And Low-Income Countries." *American Journal Of Sociology*: 1162–94.

Hill, Heather C, Brian Rowan, And Deborah Loewenberg Ball. 2005. "Effects Of Teachers' Mathematical Knowledge For Teaching On Student Achievement." *American Educational Research Journal* 42(2): 371–406.

Hinnerich, Björn Tyrefors, Erik Höglin, And Magnus Johannesson. 2011a. "Are Boys Discriminated In Swedish High Schools?" *Economics Of Education Review*.

Hinnerich, Björn Tyrefors, Erik Höglin, And Magnus Johannesson. 2011b. "Ethnic Discrimination In High School Grading: Evidence From A Field Experiment." *SSE/EFI Working Paper Series In Economics And Finance* 733.

Hirshleifer, Jack. 1985. "The Expanding Domain Of Economics." *The American Economic Review* 75(6): 53–68.

Horstschräer, Julia, And Maresa Sprietsma. 2015. "The Effects Of The Introduction Of Bachelor Degrees On College Enrollment And Dropout Rates." *Education Economics* 23(3).

Imbens, Guido W, And Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide To Practice." *Journal Of Econometrics* 142(2): 615–35.

Imbens, Guido W, And Jeffrey M Wooldridge. 2009. "Recent Developments In The Econometrics Of Program Evaluation." *Journal Of Economic Literature* 47(1): 5–86.

Jacob, Brian A., And Lars Lefgren. "Can Principals Identify Effective Teachers? Evidence On Subjective Performance Evaluation In Education." *Journal Of Labor Economics* 26.1 (2008): 101-136.

Jürges, Hendrik, Kerstin Schneider, And Felix Büchel. 2005. "The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany." *Journal Of The European Economic Association* 3(5): 1134–55.

Kane, T. J., D. O. Staiger, And J. E. Rockoff. 2010. "Searching For Effective Teachers With Imperfect Information." *Journal Of Economic Perspectives* 24(3): 97–118.

Kerr, Sari Pekkala, Tuomas Pekkarinen, And Roope Uusitalo. 2013. "School Tracking And Development Of Cognitive Skills." *Journal Of Labor Economics* 31(3): 577–602.

Koedel, Cory, Kata Mihaly, And JE Rockoff. 2015. "Value-Added Modeling: A Review." *Economics Of Education Review* (January 2015).

Kristen, Cornelia. 2003. *School Choice And Ethnic School Segregation: Primary School Selection In Germany*. Waxmann Verlag.

Krücken, Georg, Johanna Bunzmann, Lisa Hürter, Maja Kandzorra, Katharina Kloke, Juliana Körnert, Steffen Ludwig, Boris Podolsak, And Yvonne Prill. 2005. "Hochschulen Im Wettbewerb. Eine Untersuchung Am Beispiel Der Einführung Von Bachelor- Und Masterstudiengängen An Deutschen Universitäten." *Bielefeld: Department Of Sociology, University Of Bielefeld*.

Kultusministerkonferenz. 1997. "Stärkung Der Internationalen Wettbewerbsfähigkeit Des Studienstandortes Deutschland." *KMK Beschluss, Bonn*.

Kunter, M. 2011. "Motivation Als Teil Der Professionellen Kompetenz - Forschungsbefunde Zum Enthusiasmus Von Lehrkraeften." In *Professionelle Kompetenz Von Lehrkraeften: Ergebnisse Des Forschungsprogramms COACTIV*, Muenster: Waxmann, 259–75.

Kunter, Mareike, Uta Klusmann, Thamar Dubberke, Jürgen Baumert, Werner Blum, Martin Brunner, Alexander Jordan, Stefan Krauss, Katrin Löwen, Michael Neubrand, And Others. 2007. "Linking Aspects Of Teacher Competence To Their Instruction: Results From The COACTIV Project." In *Studies On The Educational Quality Of Schools: The Final Report On The DFG Priority Programme*, Waxmann, 39–59.

Ladd, H., Tim R. Sass, And Douglas N. Harris. "The Impact Of National Board Certified Teachers On Student Achievement In Florida And North Carolina." *Evidence Prepared For The National Academies Committee On The Evaluation Of The Impact Of Teacher Certification By The National Board For Professional Teaching Standards* (2007).

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' Or Boys' Human Capital Outcomes? Evidence From A Natural Experiment." *Journal Of Public Economics* 92(10-11): 2083–2105.

Lavy, Victor. 2009. "Performance Pay And Teachers' Effort, Productivity, And Grading Ethics." *The American Economic Review* 99(5): 1979–2021.

Lavy, Victor. *What Makes An Effective Teacher? Quasi-Experimental Evidence*. No. W16885. National Bureau Of Economic Research, 2011.

Levitt, Steven, Et Al. "The Behavioralist Goes To School: Leveraging Behavioral Economics To Improve Educational Performance." *Framed Field Experiments* 379 (2013).

Linden, Leigh L. 2008. *Complement Or Substitute?: The Effect Of Technology On Student Achievement In India*. Infodev.

Lochner, Lance. 2011. 4 Handbook Of The Economics Of Education *Chapter 2 - Nonproduction Benefits Of Education: Crime, Health, And Good Citizenship*. Elsevier B.V.

Lou, Yiping, Philip C Abrami, John C Spence, Catherine Poulsen, Bette Chambers, And Sylvia d'Apollonia. 1996. "Within-Class Grouping: A Meta-Analysis." *Review Of Educational Research* 66(4): 423–58.

Lucas, Robert E. 1988. "On The Mechanics Of Economic Development." *Journal Of Monetary Economics* 22(1): 3–42.

Mallar, Charles D. 1979. "Alternative Econometric Procedures For Program Evaluations: Illustrations From An Evaluation Of Job Corps." In *Proceedings Of The Business And Economics Statistics Section Of The American Statistical Association*, , 317–21.

Maluccio, John. 1998. "Endogeneity Of Schooling In The Wage Function: Evidence From The Rural Philippines." *Food Consumption And Nutrition Division Discussion Paper* 54.

Mankiw, N. Gregory, David Romer, And David N. Weil. *A Contribution To The Empirics Of Economic Growth*. No. W3541. National Bureau Of Economic Research, 1990.

Mcclellan, Mark, Barbara J Mcneil, And Joseph P Newhouse. 1994. "Does More Intensive Treatment Of Acute Myocardial Infarction In The Elderly Reduce Mortality?: Analysis Using Instrumental Variables." *Survey Of Anesthesiology* 39.3 (1995): 141.

Metzler, J., And L. Woessmann. 2012. "The Impact Of Teacher Subject Knowledge On Student Achievement: Evidence From Within-Teacher Within-Student Variation." *Journal Of Development Economics* 99(2): 486–96.

Michalos, Alex C. 2008. "Education, Happiness And Wellbeing." *Social Indicators Research* 87(3): 347–66.

Mincer, Jacob. 1958. "Investment In Human Capital And Personal Income Distribution." *The Journal Of Political Economy*: 281–302.

Mincer, Jacob. 1974. "Schooling, Experience, And Earnings. Human Behavior & Social Institutions No. 2."

Mühlenweg, Andrea M. 2010. "Teaching, Organization, And Personal Problems: Evidence From Reforming Tertiary Education In Germany." *ZEW Discussion Paper No. 10-040*.

Muralidharan, Karthik, And Venkatesh Sundararaman. *Teacher Performance Pay: Experimental Evidence From India*. No. W15323. National Bureau Of Economic Research, 2009.

Murphy, Richard, And Felix Weinhardt. "Top Of The Class: The Importance Of Rank Position." *Centre For The Economics Of Education Discussion Paper* 1241 (2014).

Nagler, Markus, Marc Piopiunik, And Martin R. West. *Weak Markets, Strong Teachers: Recession At Career Start And Teacher Effectiveness*. No. W21393. National Bureau Of Economic Research, 2015.

Nelson, Richard R, And Edmund S Phelps. 1966. "Investment In Humans, Technological Diffusion, And Economic Growth." *The American Economic Review* 56(1/2): 69–75.

NEPS. 2011a. 11 *Starting Cohort 3 - Information On The Competence Test*. Bamberg. Https://Www.Neps-Data.De/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/1-0-0/C_A28_En.Pdf.

NEPS. 2011b. 11 *Starting Cohort 4 Students - Information On The Competence Test*. Bamberg.

Oosterbeek, Hessel, Mirjam Van Praag, And Auke Ijsselstein. 2010. "The Impact Of Entrepreneurship Education On Entrepreneurship Skills And Motivation." *European Economic Review* 54(3): 442–54.

Parey, Matthias, And Fabian Waldinger. 2011. "Studying Abroad And The Effect On International Labour Market Mobility: Evidence From The Introduction Of ERASMUS." *The Economic Journal* 121(551): 194–222.

Di Pietro, G. 2012. "The Bologna Process And Widening Participation In University Education: New Evidence From Italy." *Empirica* 39(3): 357–74.

Piopiunik, Marc, Guido Schwerdt, And Ludger Woessmann. 2013. "Central School Exit Exams And Labor-Market Outcomes." *European Journal Of Political Economy* 31: 93–108.

Portela, Miguel, Carla Sa, Fernando Alexandre, And Ana Rute Cardoso. 2009. "Perceptions Of The {Bologna Process}: What Do Students' Choices Reveal?" *Higher Education* 58(4): 465–74.

Psacharopoulos, George. 1994. "Returns To Investment In Education: A Global Update." *World Development* 22(9): 1325–43.

Reiersøl, Olav. 1945. "Confluence Analysis By Means Of Instrumental Sets Of Variables."

Rivkin, S. G., E. A. Hanushek, And J. F. Kain. 2005. "Teachers, Schools, And Academic Achievement." *Econometrica* 73(2): 417–58.

Rockoff, Jonah E. 2004. "The Impact Of Individual Teachers On Student Achievement: Evidence From Panel Data." *American Economic Review*: 247–52.

Romer, Paul. *Endogenous Technological Change*. No. W3210. National Bureau Of Economic Research, 1989.

Rosenbaum, Paul R, And Donald B Rubin. 1983. "The Central Role Of The Propensity Score In Observational Studies For Causal Effects." *Biometrika* 70(1): 41–55.

Rothstein, J. 2010. "Teacher Quality In Educational Production: Tracking, Decay, And Student Achievement." *The Quarterly Journal Of Economics* (February): 175–214.

Sacerdote, Bruce. 2011. 3 Handbook Of The Economics Of Education *Chapter 4 - Peer Effects In Education: How Might They Work, How Big Are They And How Much Do We Know Thus Far?* 1st Ed. Elsevier B.V.

Sass, Tim R., Anastasia Semykina, And Douglas N. Harris. 2014. "Value-Added Models And The Measurement Of Teacher Productivity." *Economics Of Education Review* 38: 9–23.

Schlotter, Martin, Guido Schwerdt, And Ludger Woessmann. 2011. "Econometric Methods For Causal Evaluation Of Education Policies And Practices: A Non-Technical Guide." *Education Economics* 19(2): 109–37.

Schultz, Theodore W. 1961. "Investment In Human Capital." *The American Economic Review*: 1–17.

Schwerdt, G., And A. C. Wuppermann. 2011. "Is Traditional Teaching Really All That Bad? A Within-Student Between-Subject Approach." *Economics Of Education Review* 30(2): 365–79.

Schwerdt, Guido, Martin R. West, And Marcus A. Winters. *The Effects Of Test-Based Retention On Student Outcomes Over Time: Regression Discontinuity Evidence From Florida*. No. W21509. National Bureau Of Economic Research, 2015.

Seidel, T., And R. J. Shavelson. 2007. "Teaching Effectiveness Research In The Past Decade: The Role Of Theory And Research Design In Disentangling Meta-Analysis Results." *Review Of Educational Research* 77(4): 454–99.

Silles, Mary A. 2009. "The Causal Effect Of Education On Health: Evidence From The United Kingdom." *Economics Of Education Review* 28(1): 122–28.

Slavin, R. E., C. Lake, And C. Groff. 2009. "Effective Programs In Middle And High School Mathematics: A Best-Evidence Synthesis." *Review Of Educational Research* 79(2): 839–911.

Smith, Jeffrey A, And Petra E Todd. 2005. "Does Matching Overcome Lalonde's Critique Of Nonexperimental Estimators?" *Journal Of Econometrics* 125(1): 305–53.

Solow, Robert M. 1956. "A Contribution To The Theory Of Economic Growth." *The Quarterly Journal Of Economics*: 65–94.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal Of Economics* 3.

Spieß, C Katharina, And Katharina Wrohlich. 2010. "Does Distance Determine Who Attends A University In Germany?" *Economics Of Education Review* 29(3): 470–79.

Sprietsma, Maresa. 2012. "Discrimination In Grading: Experimental Evidence From Primary School Teachers." *Empirical Economics* 45(1): 523–38.

Statistisches Bundesamt. 2007. "Bildung Und Kultur. Studierende An Hochschulen." *Fachserie 11 Reihe 4.1.*

Stiglitz, Joseph E. 1999. "Knowledge As A Global Public Good." *Global Public Goods* 1(9): 308–26.

Swan, Trevor W. 1956. "Economic Growth And Capital Accumulation." *Economic Record* 32(2): 334–61.

Thistlethwaite, Donald L, And Donald T Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative To The Ex Post Facto Experiment." *Journal Of Educational Psychology* 51(6): 309.

Todd, Petra E., And Kenneth I. Wolpin. 2003. "On The Specification And Estimation Of The Production Function For Cognitive Achievement*." *The Economic Journal* 113(485):

Tyler, John H, Eric S Taylor, Thomas J Kane, And Amy L Wooten. 2010. "Using Student Performance Data To Identify Effective Classroom Practices." *The American Economic Review*: 256–60.

Vandenbussche, Jerome, Philippe Aghion, And Costas Meghir. 2006. "Growth, Distance To Frontier And Composition Of Human Capital." *Journal Of Economic Growth* 11(2): 97–127.

Voss, Thamar, Thilo Kleickmann, Mareike Kunter, And Axinja Hachfeld. 2011. "Überzeugungen Von Mathematiklehrkräften." In *Professionelle Kompetenz Von Lehrkr{Ä}Ften: Ergebnisse Des Forschungsprogramms COACTIV*, Waxmann, 235–57.

Weiss, Andrew. 1995. "Human Capital Vs. Signalling Explanations Of Wages." *The Journal Of Economic Perspectives* 9(4): 133–54.

Winter, Martin, Birgitt A Cleuvers, And Yvonne Anger. 2010. Studien Zum Deutschen Innovationssystem Studien Zum Deutschen Innovationssystem Nr. 12-2010, In *Implikationen Der Gestuften Hochschul-Curricula Auf Die Innovationsfähigkeit Deutschlands. Qualitative Untersuchungen Zur Umstellung Der Studien-Curricula In Deutschland*. Berlin: Expertenkommission Forschung Und Innovation.

Wissenschaftsrat. 2000. "Empfehlungen Zur Einführung Neuer Studienstrukturen Und -Abschlüsse (Bakkalaureus/Bachelor - Magister/Master) In Deutschland." *Drs. 4418/00, Köln*.

Woessmann, L. 2005. "Educational Production In Europe." *Economic Policy* 20(43): 446–504.

Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, Martin R West, And Others. 2009. "School Accountability, Autonomy And Choice Around The World." Edward Elgar, Cheltenham, UK.

Wright, Philip Green. 1928. "Tariff On Animal And Vegetable Oils" New York: Macmillan.

Zwick, Thomas. 2006. "The Impact Of Training Intensity On Establishment Productivity." *Industrial Relations: A Journal Of Economy And Society* 45(1): 26–46.