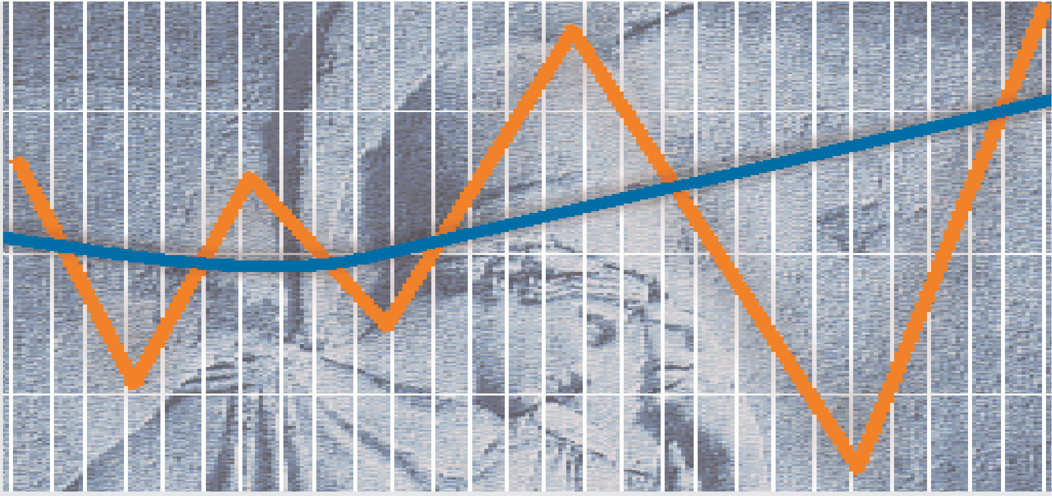


ifo Beiträge
zur Wirtschaftsforschung

50

Institutional Determinants of Student Achievement Microeconometric Evidence

Susanne Link

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Herausgeber der Reihe: Hans-Werner Sinn
Schriftleitung: Chang Woon Nam

ifo Beiträge
50 zur Wirtschaftsforschung

Institutional Determinants of Student Achievement Microeconometric Evidence

Susanne Link

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Bibliografische Information der Deutschen Nationalbibliothek

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

ISBN-13: 978-3-88512-542-6

Druck: ifo Institut, München

ifo Institut im Internet:
http://www.cesifo-group.de

Preface

This volume was prepared by Susanne Link during her stay at the Ludwig-Maximilians-University of Munich and the Department of Human Capital and Innovation of the Ifo Institute of Economic Research. It was accepted as a doctoral thesis by the Economics Departure of the University of Munich in December 2012.

The thesis consists of four core chapters, each evaluating the effect of a policy reform. Chapters 2 to 4 provide a contribution to the literature on economics of education and investigate the impact of three educational institutions on student achievement. Chapter 5 evaluates the success of the Federal Expellee Law, which was introduced to improve the economic situation of the expellees after they were forced to leave their homelands in the aftermath of World War II. The econometric analyses are based on different micro data sets and employ micro-econometric methods to identify causal effects.

Chapter 2 analyzes the effect of school autonomy on student achievement. We develop a cross-country panel using the four available PISA waves and show that school autonomy has heterogeneous effects across levels of economic and educational development. Moreover, we present evidence that local decision-making works better when there is also external accountability that limits any opportunistic behavior of schools.[1]

Chapter 3 contributes to a growing quasi-experimental literature on the effects of single-sex schooling. By exploiting the fact that students in Korea are randomly assigned to schools, we find positive, significant effects for girls with low supporting parental backgrounds in math from attending a single-sex school. Examining the underlying mechanisms suggests that parts of the effect can be attributed to a rougher classroom climate at mixed schools.

Chapter 4 investigates the effect of modified program policies on student achievement at a German university. By using a difference-in-differences approach, we find that the first reform, which effectively doubled the time until students receive their certificates, and which reduced the impact of each exam on the Grade Point Average, had a negative impact on student achievement. Furthermore, we find that a higher number of allowed

[1] This chapter was coauthored by Eric A. Hanushek and Ludger Woessmann and is accepted for publication as "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA", *Journal of Development Economics*, forthcoming, Elsevier.

resits increased the portion of students that submitted blank papers, so that they can resit the exam and improve.[2]

Chapter 5 contributes to the literature on the assimilation of migrants and studies the effect of an integration policy on expellees' labor market situation in the context of a forced mass migration. By comparing expellees to similar groups of local West Germans, we find no evidence that the law met its goal to foster the expellees' labor market integration.

Keywords: Student achievement, school autonomy, decentralization, developing countries, educational production, international student achievement tests, panel estimation, single-sex schooling, random assignment, peer effects, incentives, higher education, difference-in-differences approach, forced migration, integration policy.

JEL-Codes: D04, I20, I21, I23, I24, I25, I28, J16, J61, N30, O15

[2] This chapter was coauthored by Oliver Falck and Stephan Heblich and was published as "Forced Migration and the Effects of an Integration Policy in Post-World War II Germany", *B.E. Journal of Economic Analysis & Policy: Topics* 12 (1) 2012, De Gruyter.

Acknowledgements

I am very grateful to Ludger Woessmann, the supervisor of my thesis, for constant support, insightful discussions and valuable comments on previous versions of this thesis. I also thank Alexander Danzer for serving as the second supervisor and for valuable comments on parts of this thesis. Furthermore, I thank Amelie Wuppermann for being the third examiner of this thesis.

I sincerely thank Eric A. Hanushek, Ludger Woessmann, Philipp Beltz, Andreas Ostermaier, Oliver Falck and Stephan Heblich for the collaborations on Chapters 2, 4, and 5. I especially thank former and actual colleagues at the Department of Human Capital and Innovation at the Ifo Institute for inspiring conversations and helpful suggestions. In particular, I want to thank Nadine Fabritz, Teresa Buchen, and Erik Hornung for constant support and helpful advice. I also thank research assistants at the Department of Human Capital and Innovation for their invaluable help.

I gratefully acknowledge financial support from the German Science Foundation (DFG) through GRK 801 which gave me the opportunity to undertake the research presented in this thesis. I further thank the Munich Graduate School and the Ifo Institute for financial support which enabled me to present my research at national and international conferences.

Finally, I thank my family and friends, who supported and encouraged me throughout the whole time.

Institutional Determinants of Student Achievement

Microeconometric Evidence

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

2012 vorgelegt von

Susanne Link

Referent: Prof. Dr. Ludger Wößmann
Korreferent: Prof. Dr. Alexander Danzer
Promotionsabschlussberatung: 15. Mai 2013

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Importance of Human Capital

Nowadays, the importance of human capital investment for individual success and economic development is widely accepted. There exists compelling evidence that higher-educated individuals earn higher wages, experience less unemployment, and work in more prestigious occupations (e.g., Card, 1999; Mulligan, 1999). Moreover, recent studies confirm that human capital, in particular cognitive skills, are a key driver of long-run economic growth and development (see Hanushek and Kimko, 2000; Hanushek and Woessmann, 2008, 2011a, 2012). Consequently, much political attention is directed toward the endowment of students with basic skills and the improvement of overall educational performance.

However, there still exists a large uncertainty which education policies indeed promote the development of cognitive skills. Although education policies on school resources, such as class size reductions, receive much attention, increases in educational spending are not necessarily accompanied by achievement gains (see, e.g., Hanushek, 2003; Woessmann, 2007). By contrast, there exists compelling evidence that the institutional set-up of education systems is decisive for student achievement. It is thus important to analyze and identify institutional features that change the educational environment to increase overall educational performance.

1.2 Determinants of Educational Achievement

Substantial research has gone into understanding the determinants of educational achievement. One of the earliest studies was the Coleman Report (Coleman, 1966) that provides a detailed description of the educational achievement of about 600,000 American students, their teachers, and their schools. This report became very influential, since it directed the attention of researchers and policy-makers to the relationship between school inputs and student achievement.

As economist entered the field, this relationship became known as the "education production function", which draws an analogy between the knowledge creation process at schools and the production process at firms.[1] A standard education production function can be depicted as

$$Y_i = f(F_i, P_i, A_i, S_i). \tag{1.1}$$

This formula simply states that educational output (Y_i) of student i is a function of cumulative influences of the student's family background (F_i), of cumulative peer influences (P_i), of indiviudal ability (A_i), and of cumulative influences from school inputs (S_i) (Hanushek, 1970, 1979).

With the emergence of national and international student assessment data, this framework has been tested empirically in an enormous number of studies. The most recent assessments – such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme in International Student Assessment (PISA) – provide educational achievement indicators along with an extensive set of background information on the student, school and teacher level.[2] The observable educational outcome is then related to contemporaneous measures of relevant input factors in the empirical analysis.[3]

While theoretical considerations suggest a large number of desirable outcome as well as input measures, the availability of data has restricted empirical analyses to focus on broader categories or approximations. For example, the process of education produces several outcomes, including the development of cognitive and non-cognitive skills as well as socialization and character building. However, early studies concentrated mainly on years of schooling as a measure for human capital in both microeconometric (e.g., Psacharopoulos, 1994; Card, 1999; Psacharopoulos and Patrinos, 2004) as well as

[1] See Hanushek (1979) for background information on the development of the education production function.

[2] For an overview on international tests of educational achievement see, e.g, Hanushek and Woessmann (2011b).

[3] See Todd and Wolpin (2003) for a discussion on the specification and estimation of production functions.

macroeconometric studies (e.g., Barro, 1991; Mankiw et al., 1992; Barro and Lee, 1993). More recently, standardized achievement test scores, provided by e.g. PISA or TIMSS, are commonly used to approximate the level of cognitive development (e.g., Heyneman and Loxley, 1983; Bishop, 1997; Hanushek and Kimko, 2000), although student attitudes, college drop-out rates and time of graduation are also studied (e.g., Garibaldi et al., 2012). Most recently, the emergence of very comprehensive data sets, that include, e.g., non-cognitive skill measures, allows research to focus on even more facetted outcomes of the process of education (e.g., Dee and West, 2011).

The same is true for the inputs that enter the education production function. A child's educational development is strongly influenced by the family's support and interest in education. Although difficult to observe, these things are strongly correlated with observable measures related to a family's socio-economic status. Thus, in the empirical analysis, measures such as parents' education and occupation, family goods and family size are commonly used to account indirectly for these effects. A large number of studies document a strong association of educational achievement with many measures of family background (see, e.g., Woessmann, 2003; Schütz et al., 2008; Woessmann, 2008). Although family background is not directly malleable by policy, these estimates provide an important indication of the equality of education and the intergenerational mobility of a society.

Another set of important influences comes from a student's peers, which includes both the friends outside the school and the class- and schoolmates.[4] Not only a student's motivation and attitude toward learning, but also classroom atmosphere and teaching style are likely to be influenced by the character of the student body. The literature on peer effects is based on the belief that students perform better if they are educated amongst better school- and classmates. A method to account for peer characteristics in the empirical analysis is to include aggregate measures of family background and test scores in the regression. While there is some evidence for the existence of peer effects, they are mostly small in magnitude (Sacerdote, 2001; Hanushek et al., 2003; Angrist and Lang, 2004; Ammermüller and Pischke, 2009).[5]

Individual ability is another important determinant of educational achievement that needs to be considered in the conceptual framework of the education production process.

[4] See Epple and Romano (2011) for a survey on the theory and evidence of peer effects.

[5] Even though naive estimations show positive peer effects, the literature is mostly concerned with the separation of social effects from other confounding influences (Manski, 1993).

Although the concept of “ability” is not well defined in the literature, it refers to any innate and acquired endowments or learning capacities that facilitate the accumulation of knowledge and skills (Hanushek, 1979). While individual ability is crucial to the conceptual model of learning, it is most often neglected in the empirical analysis due to data constraints. More recently, some studies approximate ability by controlling for IQ scores or try to account for unobserved ability by using microeconometric methods (see Chapter 1.3).

On the school level, factors such as class size, school facilities, and teacher characteristics are potential determinants for student achievement, and often available in student assessment data sets. Since improvements at school present a natural starting point for policy interventions, the most extensive literature focus on the effects of school resources. However, according to the existing literature, policies that involve increases in educational spending do not necessarily improve student performance (see, e.g., Hanushek, 2003; Woessmann, 2007). The quality of teaching presents another key determinant of school quality. Although the literature on teacher characteristics shows that observed teacher characteristics, including education and experience, do not account for much of the variation in student test scores, large differences in student learning across teachers support the idea that teacher quality matters (see, e.g., Hanushek, 1971; Rockoff, 2004; Rivkin et al., 2005).

To design education policies, it is important to identify determinants of student learning which can be controlled by policy. In contrast to family background and individual ability, school resources and teacher quality are generally malleable by policy. Based on the existing evidence, however, it is difficult to formulate specific education policies. Thus, research has increasingly focused on the larger institutional issues of an education system and the incentives for the people involved in the education process, namely students, teachers and principals. This development is supported by standard economic theory that predicts that persistent institutional changes will alter incentives and thus behavior (Hanushek and Woessmann, 2011b).

The basic model underlying this literature is the education production function which is extended by the institutional features of schools and education systems (I_i) (Hanushek and Woessmann, 2011b):

$$Y_i = f(F_i, P_i, A_i, S_i, I_i). \tag{1.2}$$

So far, five institutional features have received particular attention in the international literature. The existing evidence suggests that accountability in terms of central exit exams, autonomy (in combination with accountability), and competition from private schools are associated with higher student performance (see, e.g., Bishop, 1997; Woessmann et al., 2009; West and Woessmann, 2010). Later tracking and compulsory pre-primary education seems to be beneficial for the equality of student outcomes (see, e.g., Hanushek and Woessmann, 2006; Woessmann et al., 2009; Schneeweis, 2010).[6]

Chapters 2 to 4 of this thesis contribute to the literature on educational institutions and investigate three aspects of education systems empirically. In particular, Chapter 2 adds to the existing evidence on the effects of school autonomy and investigates whether effects vary with a country's level of development. Chapter 3 and 4 are concerned with the organization of learning. Chapter 3 analyzes the effect of single-sex schooling on student achievement, Chapter 4 examines the effect of program rules on the achievement of university students. In a similar vein, Chapter 5 evaluates the success of a labor market policy. In the following, the microeconometric methods used in this thesis are outlined and a summary of Chapters 2 to 5 is presented.

1.3 Causal Inferences in Economics of Education

Measuring educational performance and understanding its determinants is important to design education policies. In particular, the identification of causal relationships is crucial to provide policy recommendations. However, estimating causal effects presents a major challenge within the empirical literature. Although the education production function has been investigated empirically in an enormous number of studies, most of this work relates the observable achievement indicators to information on student, family and school characteristics in simple regression analyses. Unfortunately, there are a number of reasons that might bias the estimated coefficients.

---

[6] See, e.g., Hanushek and Woessmann (2011b) for a literature review on the effects of accountability, autonomy, competition from private schools, tracking and pre-primary education.

A natural starting point to evaluate the effects of a particular institution is to compare observational units, e.g. schools or countries, with different institutions in a regression analysis. However, even if students at school A outperform students at school B, this is not necessarily *caused* by school A's institutions. In fact, there are a number of reasons that might explain this particular association. For example, the effect of attending a private school cannot be inferred by comparing student achievement at private and public schools, because students with well supporting parental backgrounds are both, more likely to sort into private institutions and more likely to receive high results. Similarly, cross-country comparisons have their difficulties. Although international achievement data makes the analysis of systematic variation in institutions across countries possible, again threats to causal inference arise. This is easily seen, as there are not only differences in the institution of interest, but in culture, other institutions, values and so forth that might be correlated with the particular institution and also affect the outcome of interest.

A first attempt to overcome these so-called endogeneity problems is to account for confounding factors by including a large set of control variables into the model. However, if there are any unobserved determinants of the outcome variable that are also correlated with the variables of interest, the estimated coefficients are still biased. One obvious example is individual ability, a major determinant of educational success, that is usually unobserved, but likely to be correlated with individual decisions and characteristics (see Card, 1999). Another example are cultural differences in cross-country studies that are difficult to measure, but likely to be influential in a number of ways.

Panel data methods are one way to control for important forms of unobservable heterogeneities. By observing the same unit of observation, e.g., a student, school, or country, at several points of time, it is possible to account for time-invariant, unit-specific heterogeneities by employing fixed-effects models. The coefficient of interest is then estimated by exploiting the variation in the variable of interest over time within the same unit of observation.[7] For example, a cross-country panel can be used to evaluate the effect of a particular institution while accounting for unobserved, time-invariant cross-country differences (see Chapter 2). The underlying identifying assumption is that there are no

[7] This idea is also implemented in cross-sectional data. For example, observing the performance of the same individual across different subjects allows to account for any factors that affect different subjects in the same way, such as general retentiveness or motivation (see, e.g., Metzler and Woessmann, 2012).

unobserved, influential time-varying factors that are correlated with the variation in the institutional setting.

The difference-in-differences approach (Campbell, 1969; Card and Sullivan, 1988) is a simple panel data method that is commonly applied to situations where a policy reform affects only a particular group (treatment group) while leaving the situation for a nearly identical group (control group) unchanged (see Chapter 4 and 5). In the simplest case, both groups are observed before and after the implementation of the reform. By comparing the change in the outcome variable between treatment and control group, the difference-in-differences estimator identifies a causal treatment effect while accounting for time-invariant unobservable differences at the group-level. The identifying assumption is that both groups would have developed with the same trend in the absence of the treatment such that the development of the control group can be used as a counterfactual scenario for the development of the treatment group.

While panel data methods account for unobserved time-invariant heterogeneities, the identification of causal effects requires further assumptions. These assumptions need to be verified and can sometimes be tested empirically. In the cross-country panel example, one needs to argue that the observed institutional change is not driven by other factors, such as initially poor performance, to rule out a reverse causality or omitted variable problem. Crucial to the difference-in-differences approach is – in addition to the common trend assumption – that the composition of the treatment and control groups does not change as a result of the treatment. Overall, these assumptions address the remaining threat that the variable of interest is endogenously related to the outcome variable.

The most convincing way to obtain causal effects is therefore to exploit exogenous variation in the variable of interest. Ideally, one would like to run a controlled experiment where treatment and control status, or – termed in the previous spirit – institutional status, are randomly assigned to draw causal inferences. This ensures that the variable of interest is independent of the potential outcomes (Angrist and Pischke, 2008). Although controlled field experiments are a growing phenomenon for policy evaluations, especially in the U.S. and in many developing countries, randomized trials involve high costs and long durations. Therefore, researchers have developed ways to evaluate reforms by using observational data in an experimental spirit.[8] The underlying idea is to find so-called natural experiments, a term that refers to situations in which there is a transparent exogenous source of variation in the variable of interest. This variation may be induced by government randomization, or other events that involve unforeseeable changes of the conditions of interest (Meyer, 1995).

For example, assignment on oversubscribed programs is often handled by randomized lotteries, so as to give each applicant an even chance for participation (for an overview see Schlotter et al., 2011). Since the treatment has been randomly assigned among those who applied, the comparison between the lottery winners and the lottery losers then identifies the causal program effect (see, e.g., Peterson et al., 2003). Another reason for government randomization are aspects of equity. In South Korea, for example, students are randomly assigned to schools to avoid the clustering of students with high socio-economic backgrounds at privileged schools. This ensures that attendance at a particular school is independent of family background and unobserved student heterogeneities which helps to identify causal effects (see Chapter 3).

1.4 Outline of This Thesis

This thesis contributes to the literature on the determinants of student achievement. As the existing literature has identified institutional features of schools and education systems as highly influential, the role of three popular institutions on student achievement will be investigated. Conceptually, the analyses in Chapters 2 to 4 are based on the education production function that directly relates educational achievement to the relevant input factors at the student, school, and institutional level (Section 1.2). Methodologically, microeconometric methods are employed to address the need for causal inference in policy-evaluations (Section 1.3). In a similar approach, the success of a labor market policy aimed at the integration of expellees is evaluated in Chapter 5. Chapter 6 concludes with a summary of the key findings developed in this thesis and derives important policy implications. In the following, an outline of Chapters 2 to 5 is presented.

---

[8] See, e.g., Angrist and Pischke (2008) for an overview on microeconometric methods to identify causal effects. See, e.g., Schlotter et al. (2011) for a non-technical guide on microeconometric methods for causal evaluation of education policies.

Chapter 2 analyzes the effect of school autonomy on student achievement.[9] The term "school autonomy" refers to the devolution of decision-making powers from education authorities to schools, for areas such as hiring of teachers or the choice of curricular elements. Although the policy of school autonomy became very popular within the last decades, there have been opposing movements internationally. While some countries gave more decision-making power to schools, other countries have actually implemented more centralization. These opposing movements reflect the underlying tension associated with the effects of autonomy. Arguments favoring school autonomy include the fact that teachers and principals have superior local knowledge, which is needed to make efficient decisions and to respond to students' needs. However, given incomplete information, school autonomy also opens up the possibility of opportunistic behavior. Moreover, even if interests are aligned, the lack of decision-making capacity may lead to poor decisions. Given these opposing mechanisms, Chapter 2 develops the idea that school autonomy may have heterogeneous effects across levels of economic and educational development, and across different regimes of centralized accountability. In particular, school autonomy is likely to be interacting with a country's level of development that captures elements such as strong institutions, governmental effectiveness, informal accountability mechanisms and decision-making capacity. Thus, a well-developed environment presents the prerequisite for beneficial effects to unfold, or, respectively, school autonomy may be even harmful in a less favorable setting. We develop a cross-country panel using the four available PISA waves of 2000, 2003, 2006, and 2009 and test if the effect of school autonomy varies with a country's level of development. Although we estimate our micro models with over one million student observations to account for family and school inputs at the individual level, the panel character of the analysis is at the country level. Thus, we identify the effect of school autonomy by exploiting within country variation over time. By using country fixed effects, we are able to control for time-invariant differences across countries such as culture and attitude. We find that school autonomy, most pronounced academic-content autonomy, has positive effects for developed countries, but may actually be harmful in less developed countries. Moreover, there is evidence that local decision-making works better when there is also external accountability that limits any opportunistic behavior of schools.

[9] This chapter was coauthored by Eric A. Hanushek and Ludger Woessmann and is accepted for publication as "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA", *Journal of Development Economics*, forthcoming, Elsevier.

Further, having generally well-functioning schools, indicated by initial performance levels, appears complementary with autonomy.

Chapter 3 analyzes the effect of single-sex schooling on student academic achievement. Although at present very few countries educate their students in gender-segregated classes or schools, there is a constant interest in whether single-sex schooling might have beneficial or adverse effects for boys and girls. One argument favoring single-sex schooling is the idea that students are distracted by the opposite sex at school and that both boys and girls may benefit from gender separation (see, e.g., Coleman, 1961). Moreover, the fact that girls and women are still underrepresented in stereotypically male subjects and occupations has stimulated the debate whether single-sex schooling may help to reduce gender stereotypes. Despite constant attention, the evidence on the effect of single-sex schooling is inconclusive. While several studies suggest positive effects, especially for girls, on educational outcomes, the problem of selection into single-sex schools is often not addressed in these studies. Chapter 3 contributes to a growing quasi-experimental literature and investigates the effect of single-sex schooling in a particular interesting setting. In Korea, students are randomly assigned to schools, regardless whether they are single-sex or coeducational organized. This ensures that attendance at single-sex schools is not correlated with unobservable individual characteristics that also influence achievement. Thus, the comparison of girls (boys) at single-sex schools and girls (boys) at coeducational schools should identify a reliable estimate of the effects of single-sex schooling on student achievement. Moreover, the rich dataset used in this study allows us to explore the underlying mechanism and channels. We find positive, significant effects for girls with low supporting parental backgrounds in math from attending a single-sex school. In contrast, there are neither beneficial nor adverse effects for boys. Although arguments favoring single-sex schooling often include differences in teaching style and student attitude, the positive effects for girls can neither be explained by differences in school and teacher characteristics nor by gender-tailored teaching practices or more positive attitudes toward math at single-sex schools. However, parts of the effect can be attributed to a rougher classroom climate at mixed schools.

Chapter 4 investigates the effect of program policies on student achievement. Our findings suggest that university program rules, such as credit points or the number of resits students can take, serve as incentives. The existing literature on incentives and academic performance mostly focuses on the effects of monetary rewards (Angrist and Lavy, 2009;

Leuven et al., 2010; Garibaldi et al., 2012). In contrast, program rules are generally inexpensive and must be adopted when universities design programs. In particular, we consider a business school at a German university that offers two similar study programs that both became subject to reforms. While the policies for the first program were changed as early as 2005, the reform of the second program was delayed until 2010. To analyze the effects of the modified reforms, we digitized students' performance data in a course that is compulsory for both groups of students and merged them with personal characteristics. By using a difference-in-differences approach, we find that the first reform, which effectively doubled the time until students receive their certificates, and which reduced the impact of each exam on the Grade Point Average, had a negative impact on student achievement. Furthermore, we find that a higher number of allowed resits increased the portion of students that submitted blank papers, so that they can resit the exam and improve. We also show that students respond differently to university policies depending on their ability. The fact that both groups of students attend the same course, are taught by the same instructors, use the same textbooks and teaching materials, and that their curricula are nearly identical when they take the exam corroborates the common trend assumption that we need to make.

Chapter 5 analyzes a policy change in a related field of research and contributes to the literature on the assimilation of migrants.[10] In particular, we study the effect of an integration policy on expellees' labor market situation in the context of a forced mass migration. After World War II, significant territorial changes forced 8 million of ethnic Germans to leave their homelands in East Prussia, Silesia, Pomerania, and Bohemia and settle within the new borders of West Germany (cf. Schmidt, 1994). After their displacement, many expellees experienced a huge loss in status. While many of them owned real estate or were self-employed before World War II, large fractions of expellees became occupied in low skilled jobs or even unemployed. As a response, the German government introduced the Federal Expellee Law (Bundesvertriebenengesetz) in 1953 with the goal of restoring the expellees' status and improving their situation. To evaluate the success of this law, we use data from the 1971 micro census that allow us to identify and distinguish expellees from local West Germans. We especially benefit from an extension

[10] This chapter was coauthored by Oliver Falck and Stephan Heblich and was published as "Forced Migration and the Effects of an Integration Policy in Post-World War II Germany", *B.E. Journal of Economic Analysis & Policy: Topics* 12 (1) 2012, De Gruyter.

of the 1971 census that was designed to gain insight into expellees' integration into the German labor market and society and contains detailed retrospective information on the occupation of the German population. By comparing expellees to similar groups of local West Germans, we find no evidence that the law met its goal to foster the expellees' labor market integration. We therefore conclude that the improved economic situation of expellees can be attributed to the general economic boom in the aftermath of World War II and not to the provision of the Federal Expellee Law.

Chapter 2

Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA*

2.1 Introduction

Virtually every country in the world accepts the importance of human capital investment as an element of economic development, but this has introduced a set of important policy questions about how best to pursue such investments. Over time, attention has shifted away from simply ensuring access to schooling to an interest in the quality of learning.[1] This shift has introduced new policy uncertainty since the process of expanding school attainment is better understood than is the process of improving achievement, leaving many countries with limited success after adopting a variety of popular policies. The uncertainty has perhaps been largest in the case of institutional design questions, as the evidence in that

---

* This chapter was coauthored by Eric A. Hanushek and Ludger Woessmann and is accepted for publication as "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA", *Journal of Development Economics*, forthcoming, Elsevier.

[1] Hanushek and Woessmann (2008) show that cognitive skills can have substantial impacts on economic development. At the same time, access and attainment goals dominate many policy discussions. The clearest statement of school attainment goals can be found in discussions of the Education for All Initiative of the World Bank and UNESCO (see the description in http://en.wikipedia.org/wiki/Education_For_All, accessed July 31, 2011) and the Millennium Development Goals of the United Nations (see the description in http://en.wikipedia.org/wiki/Millennium_Development_Goals, accessed July 31, 2011). In both instances, while there is some discussion of quality issues, the main objective has been seen as providing all children with at least a lower secondary education.

area has been thinner and less reliable. This chapter focuses on one popular institutional change – altering the degree of local school autonomy in decision-making – and brings a new analytical approach to the analysis of its impact.[2] By introducing cross-country panel analysis, we can exploit the substantial international variation in policy initiatives focused on autonomy while controlling for the large cross-country differences in cultural and institutional factors. We find that autonomy does appear significantly to affect the performance of a country's schools, but the observed impact is quite heterogeneous across stages of development: The effect of school autonomy in decision-making is positive in developed countries, but in fact turns negative in developing countries.

Local autonomy has been a policy discussed intensively in both developing and developed countries. While many countries have moved toward more decentralization in such areas as the hiring of teachers or the choice of curricular elements, others have actually gone to more centralized decision-making. The opposing movements reflect a fundamental tension. The prime argument favoring decentralization is that local decision-makers have better understanding of the capacity of their schools and the demands that are placed on them by varying student populations. This knowledge in turn permits them to make better resource decisions, to improve the productivity of the schools, and to meet the varying demands of their local constituents. Yet, countervailing arguments, centered on lack of decision-making capacity and conflicting incentives, push in the opposite direction. With local autonomy comes the possibility that individual schools pursue goals other than achievement maximization and a potential threat to maintaining common standards across the nation. Despite these competing arguments, there remains considerable policy support for further local autonomy in decision-making (e.g., Governor's Committee on Educational Excellence, 2007; Ouchi, 2003; World Bank, 2004).

From an analytical viewpoint, four significant issues arise when trying to estimate the effect of autonomy. First, the very concept of local decision-making and local autonomy is multifaceted and difficult to measure on a consistent basis. It is possible, for example, for local schools to decide some things – such as teacher hiring or facility upgrades – and not others such as the appropriate outcome standards or the pay of teachers. Conceptually, some decisions are more appropriately made locally – e.g., operational decisions like hiring

[2] Local autonomy for decision-making is referred to in various ways including decentralized decision-making and site-based or school-based management. Here, we typically use the term local or school autonomy, although we think of it as a synonym for these alternative names.

and budget allocations where local knowledge is needed and standardization is not crucial — than others where standardization may be more desirable – e.g., course offerings and requirements (see Bishop and Woessmann, 2004).

Second, the impact of autonomy may well vary with other elements of the system. For example, local autonomy permits using localized knowledge to improve performance, but it also opens up the possibility for more opportunistic behavior on the part of local school personnel. As a result, the impact on student outcomes may well interact with the level of accountability, because centralized accountability provides a way of monitoring local behavior (Woessmann, 2005).[3] In a larger sense, the results of autonomy may depend on the performance level (Mourshed et al., 2010) and – as a corollary – on the overall development level of the country and the entire school system.

Third, much of the evidence on autonomy comes from cross-sectional analyses where any effects are not well identified.[4] Specifically, one must often question whether observable characteristics adequately describe differences in schools that are and are not granted more autonomy in decision-making. For example, if more dynamic schools get greater autonomy or if demanding parents choose autonomous schools, it is difficult to extract the independent effect of local decision-making on student achievement.

Fourth, many aspects of the locus of decision-making are set at the national level. For example, many countries set national educational standards, national assessments and accountability regimes, and various rules about what decisions are permissible at the local level, leaving little to no within-country variation in decision-making authority. Relatedly, any general-equilibrium effects are extremely difficult to disentangle if, for example, the pattern of local decision-making brings a competitive response from schools without local decision-making or if the nature of local decisions alters the supply of teachers or administrators. But dealing with these issues through international comparisons – where

[3] Such considerations have also entered into the interpretation of mixed results from autonomy in the U.S. (see Hanushek, 1994; Loeb and Strunk, 2007). A further U.S. example comes from charter schools, which depend significantly on the regulatory environment they face. Charter schools are publicly financed and regulated schools that are allowed to have considerable autonomy, frequently being stand-alone schools. At least a portion of the variation in the evaluations of charter schools probably reflects interactions with other forces such as degree of parental choice, the quality of information, and constraints on school location. For estimates of the variation in charter outcomes, see CREDO (2009); Hanushek et al. (2007); Booker et al. (2007), and Bifulco and Ladd (2006).

[4] Note that more recent investigations, particularly in developing countries, have relied on randomized control trials – although these are difficult to implement and a number have not been well executed (Patrinos, 2011). There has also been more attention to evaluations built around natural experiments; see Galiani and Perez-Truglia (2011).

institutional variation can be found – brings other identification issues related to variations in culture, governmental institutions, and other things that are difficult to measure.

This chapter introduces new international panel data to shed light on each of these issues. We develop a panel of international test results from the Programme for International Student Assessment (PISA), covering 42 countries and four waves that span a time period of ten years.[5] Although we estimate our micro models with over one million student observations to account for family and school inputs at the individual level, the panel character of the analysis is at the country level. The survey information that accompanies the student assessments provides rich detail about individual students and schools along with specific descriptions of the decisions that are and are not permissible at the school level.

As pointed out, identification of the influence of specific institutional features of educational systems is obviously challenging. We directly address the most significant threats to identification of the effects of autonomous decision-making, but we cannot be sure that we have eliminated all potential problems. To begin with, by aggregating to the country level, we ensure that our estimates are not affected by within-country selection into autonomy. At the country level, however, we must deal with the myriad of ways that countries and their school systems differ. All prior cross-country work on these questions has been purely cross-sectional, necessitating strong assumptions about the adequacy of controls for systematic country-specific heterogeneity (see Hanushek and Woessmann, 2011b).[6] With the panel data on school performance, however, we can exploit country-level variation in school autonomy over time while including country (and year) fixed effects to control for systematic, time-invariant cultural and institutional differences at the country level in a very general way.[7] Within this fixed-effect framework, we can readily test the heterogeneous effects of autonomy across specific types of decisions; across variations in development levels and educational performance levels; and across different regimes of centralized accountability. Our central finding is that local autonomy has an

[5] For a discussion of international assessments along with background material for this analysis, see Hanushek and Woessmann (2011b).

[6] For examples of existing investigations of institutions – and particularly of autonomy – across countries, see Woessmann (2003, 2005); Fuchs and Woessmann (2007); Woessmann et al. (2009).

[7] An early discussion of the underlying concept can be found in Gustafsson (2006). Brunello and Rocco (2011) is a rare exception using the PISA data as a panel with country fixed effects, albeit using only country-level data, to estimate effects of the share of immigrant students on natives' test scores.

important impact on student achievement, but this impact varies systematically across countries, depending on the level of economic and educational development. In simplest terms, countries with otherwise strong institutions gain considerably from decentralized decision-making in their schools, while countries that lack such a strong existing structure may actually be hurt by decentralizing decision-making. The negative effect in developing countries emerges most clearly for autonomy in areas relating to academic content, but also appears for autonomy in the areas of personnel and budgets. An extensive series of robustness and specification tests corroborates the central finding.

We primarily use the income level of a country (GDP per capita) as an indicator of overall skills and institutions. Higher-income countries tend to have better societal and economic institutions that promote productivity, societal vision, and smooth social interactions. As such, this indicator is broad and multifaceted, leading us also to investigate more specific and nuanced aspects of institutions. We find indications that the development of the educational system (measured by higher achievement) adds another significant dimension to the success of greater local autonomy. Further, consistent with the underlying motivation for constraining opportunistic behavior, the benefits of greater autonomy are enhanced by accountability through centralized examinations.

At a methodological level, the results show the potential perils of cross-country analyses that cannot control for other institutional and development factors. In our specific analysis, we find different and conflicting results between simple cross-sectional analysis (albeit with extensive controls of measured family and schooling inputs) and our new panel estimators. Further, the heterogeneity of results across different levels of development suggests caution in attempting to generalize from developed-country analyses to developing countries (and vice versa).

Our cross-country results also rationalize the pattern of outcomes that emerges from existing within-country studies on school autonomy. Patrinos (2011) and Galiani and Perez-Truglia (2011) provide thoughtful reviews of decentralized decision-making in developing countries, including important discussions of how a clear focus on identification (such as the use of random control trials or various instrumental-variable applications), while currently limited, influences program evaluations. But their reviews make it clear in general terms that the developing-country evidence on the effects of school autonomy is mixed at best and may even be more negative (e.g., Madeira, 2007). Thus, in their comparative review of the literature, Arcia et al. (2011) conclude that "the empirical evidence from Latin America shows very few cases in which SBM [school based management] has made a significant difference in learning outcomes (Patrinos, 2011), while in Europe there is substantial evidence showing a positive impact of school autonomy on learning (Eurydice, 2007)." Indeed, two recent studies of developed countries that pay particular attention to identification – Barankay and Lockwood (2007) for Switzerland and Clark (2009) for the United Kingdom – find substantial positive effects of local autonomy.[8] And when positive effects are found for specific decentralization programs in developing countries, they tend to be either restricted to schools located in non-poor municipalities (Galiani et al., 2008) or originate from more comprehensive school reform programs that simultaneously raised accountability from local communities (e.g., Gertler et al., 2012; Gunnarsson et al., 2009; Jimenez and Sawada, 1999). These aspects are consistent with our main theme that autonomy effects depend on the development of the socio-economic and institutional environment.

The next section discusses the underlying conceptual framework. Section 2.3 describes the new database and key variation across countries in various kinds of local autonomy. Section 2.4 develops our empirical model. Section 2.5 presents our estimation results and extensive robustness and specification tests. Section 2.6 expands the investigation of interactions to centralized examinations and the performance level of the education system. Section 2.7 concludes.

2.2 Conceptual Framework

A variety of theoretical models highlight aspects of the delegation of authority to different levels of decision-makers. In terms of public entities, the relevant work can be traced back to Oates (1972, 1999) on fiscal federalism.[9] This analysis has been expanded to consider different objectives by decision-makers at different levels, often in terms of general principal-agent models. In such models, school autonomy or the decentralization of decision-making power is framed as the delegation of a task by a principal (the government

[8] In analyzing governance aspects at the level of tertiary education, Aghion et al. (2010) show that autonomy is positively related to universities' research output in the U.S. and in Europe and argue for benefits from combining autonomy with accountability.

[9] For recent analysis, see Blöchliger and Vammalle (2012).

agency in charge of the school system), who wishes to facilitate the provision of knowledge, to agents, namely the schools (see Woessmann, 2005; Galiani et al., 2008; Barrera-Osorio et al., 2009). In the absence of divergent interests or asymmetric information, agents can be expected to behave in conformity with system objectives and greater autonomy can lead to increased efficiency of public schools (e.g., Hoxby, 1999; Nechyba, 2003), because autonomy offers the possibility of using superior local knowledge. Additionally, by bringing decisions closer to the interested local community, decentralization may improve the monitoring of teachers and schools by parents and local communities (see Galiani et al., 2008, and the references therein).

However, when divergent interests and asymmetric information are present in a decision-making area, agents have incentives and perhaps substantial opportunities to act in their own self-interest with little risk that such behavior will be noticed and sanctioned. In this case, autonomy opens the scope for opportunistic behavior, with negative consequences for outcomes (Woessmann, 2005). Agents may use their greater autonomy to further goals other than advancing student achievement. Further, the quality of decision-making may also be inferior at the local level when the technical capabilities of local decision-makers to provide high-quality services are limited and when local communities lack the ability to ensure high-quality services (see Galiani et al., 2008). Consequently, the success of autonomy reforms may depend on the general level of human capital which affects the quality of parental monitoring.[10]

Substantial empirical research has gone into understanding the impact of decentralized decision-making, but, given the variety of theoretical trade-offs, virtually none has attempted to estimate the underlying structure identified in the theoretical models. Rather, the empirical work has more modestly attempted to estimate the reduced form relationship that indicates the overall impact of decentralization on educational outcomes.

One strand of empirical work has applied rigorous micro-evaluation techniques including randomized control trials, difference-in-differences techniques, and regression-discontinuity designs in order to understand the results of specific interventions. Unfortunately, there have to date been only a small number of such rigorous studies, and they have yielded mixed results (see reviews in Barrera-Osorio et al., 2009; Galiani and Perez-Truglia,

[10] While we focus on issues of decision-making, there may also be technological differences. Centralization opens the possibility to exploit economies of scale, for example in evaluation, teacher training systems, and the like.

2011). A second strand builds on the larger body of empirical work that generally fits under the label of educational production functions and that motivates this analysis. This general production function approach has been followed in a wide range of studies designed to understand how such factors as school resources and family background affect achievement.[11] Here we take an expanded view of this approach that highlights the importance of institutions and, in particular, of local autonomy.

A typical formulation of an educational production function has student outcomes (T) as a function of family (F) and schools (S):

$$T = f(F, S) \tag{2.1}$$

Here, however, we introduce the simple idea that the productivity of any input is directly related to the institutional structure of country $c(I_c)$ that determines the basic environment and rules of schools, how decisions are made, the overall incentives in the system, and so forth:

$$T = I_c f(F, S) \tag{2.2}$$

For many analyses of educational production within countries, the institutional structure is constant, and analyses that ignore it provide accurate information about the impacts of resources even if these might not transfer well to institutional structures in other countries. In many ways, I_c is similar to total factor productivity in a macro context where it determines the efficiency with which any given set of inputs is translated into student achievement. In this formulation, we are specifically interested in investigating the decision-making institutions of different countries.

Against the background of the opposing sets of mechanisms of how autonomy affects performance, we argue that the impact of autonomy likely depends on the level of development. This is a natural extension of the micro-level evaluations of interventions within countries, where autonomy has been found to widen the distribution of outcomes because of differential impacts related to the socio-economic backgrounds of families (Galiani and Perez-Truglia, 2011). It is also consistent with the comparative review of

[11] See Hanushek (2002, 2003) on the general framework and U.S. evidence; see Woessmann (2003) on international evidence.

the literature by Arcia et al. (2011) that finds few cases of positive effects of school-based management reforms in Latin America but substantial positive evidence in Europe.

In terms of our modeling, the hypothesis is that a country's development level captures such aspects as local capacity, abilities of local decision-makers, governance effectiveness, state capacity, parental human capital, and monitoring abilities of local communities. Also specifically in the education system, systems that already work at a high performance level may have such features as external evaluations and well-trained teachers that facilitate local decision-making by setting and ensuring high educational standards.[12] In particular, either accountability systems or better parental oversight may limit the extent to which local decision-makers can act opportunistically without getting caught (Woessmann, 2005; Barrera-Osorio et al., 2009). In sum, there are a number of channels through which a higher level of development, both in the education system and in society more generally, strengthens the positive mechanisms of autonomy and weakens the negative ones.

Finally, the impact of local autonomy may differ by area of decision-making. While standardization may be important in decisions on academic content, it may not be as important in decisions on process operations and personnel-management (Bishop and Woessmann, 2004). Thus, local decision-making over basic issues of standards such as course offerings or course content might have a negative effect of autonomy when the whole system is dysfunctional. But even in such a system, local decision-making over hiring teachers and budget allocations may not be as negative.

2.3 International Panel Data

An essential component of our analytical strategy, described below, is the construction of a cross-country panel of student achievement data. For this, we can take advantage of the recent expansion of international assessments (cf. Hanushek and Woessmann, 2011b).

[12] For example, in diagnosing what leads to improved performance at different stages of development, Mourshed et al. (2010) observe that going from 'great to excellent' is such that "the interventions of this stage move the locus of improvement from the center to the schools themselves" (p. 26).

2.3.1 Building a PISA Panel Database

Our empirical analysis relies on the Programme for International Student Assessment (PISA), an internationally standardized assessment conducted by the Organisation for Economic Co-operation and Development (OECD). The PISA study, first conducted in 2000, is designed to obtain internationally comparable data on the educational achievement of 15-year-old students in math, science, and reading.

Four distinct assessments have been carried out: in 2000/2002, 2003, 2006, and 2009. In PISA 2000, 32 countries, including 28 OECD countries, participated in the assessment. In 2002, a further 11 non-OECD countries administered the PISA 2000 assessment. By PISA 2009, the latest assessment available for this study, the number of participating countries reached 65 countries including a range of emerging economies.

PISA's target population is the 15-year-old students in each country, regardless of the institution and grade they currently attend. The PISA sampling procedure ensures that a representative sample of the target population is tested in each country. Most countries employ a two-stage sampling technique. The first stage draws a random sample of schools in which 15-year-old students are enrolled, where the probability of a school to be selected is proportional to its size as measured by the estimated number of 15-year-old students attending. The second stage randomly samples 35 students of the 15-year-old students in each of these schools, with each 15-year-old student having the same sampling probability.

The performance tests are paper and pencil tests, lasting up to two hours for each student. The PISA tests are constructed to test a range of relevant skills and competencies. Each subject is tested using a broad sample of tasks with differing levels of difficulty to represent a comprehensive indicator of the continuum of students' abilities. The performance in each domain is mapped on a scale with a mean of 500 test-score points and a standard deviation of 100 test-score points across the OECD countries.[13]

PISA makes a concerted effort to obtain random samples of the school population and to monitor the testing conditions. In fact, when conditions do not meet the standards, a

[13] While the reading test has been psychometrically scaled on a uniform scale since 2000, the math test was re-scaled in 2003 (and the science test in 2006) to have again mean 500 and standard deviation 100 across the OECD countries and has a common psychometric scale since then. In our analyses below, year fixed effects take account of this. Furthermore, we show that results are qualitatively the same when restricting the math analysis to the waves since 2003 that have a common psychometric scale.

country's results are not reported.[14] For some developing countries, a number of students have dropped out by age 15, which could bias the testing. The impact of this potential problem is tested in the robustness section below.

In addition to the achievement data, PISA also provides a rich array of background information on each student and her school. Students are asked to provide information on personal characteristics and their family background. School principals provide information on the schools' resource endowment and institutional settings. While some questionnaire items, such as the questions on student gender and age, remain the same in each assessment cycle, some information is not available or directly comparable across all PISA waves.

By merging the four PISA assessment cycles, we are, for the first time, able to construct a panel dataset at the country level. In a first step, we combine students' test scores in math, science, and reading literacy with individual students' characteristics, family background information, and school-level data for each of the four PISA waves. Since the background questionnaires are not fully standardized, in a second step we select a set of core variables that are available in each of the four PISA waves and merge the cross-sectional data of 2000/2002, 2003, 2006, and 2009 into one dataset.

Our sample comprises all countries that participated in at least three of the four PISA waves.[15] Combining the available data, we construct a dataset containing 1,042,995 students in 42 countries. As is evident from Table 2.1, the panel includes a broad sample of both high-income and lower-income countries. Following the World Bank classification, 25 countries in our sample are classified as high-income countries. But there is also one low-income country, seven lower-middle-income countries, and nine upper-middle-income countries in the sample.

Figure 2.1 depicts the available achievement data for the 42 countries in our sample. The average test performance across all countries in the sample hardly changed between 2000 and 2009 (see also Table 2.1). But some countries saw substantial increases in average achievement (most notably Brazil, Luxembourg, Chile, Portugal, Mexico, and Germany

---

[14] For example, because of deviations from protocol, the United Kingdom scores were not reported in 2003, the scores for the Netherlands in 2000, and the U.S. reading scores in 2006. While the United Kingdom scores for 2000 are included in the database, subsequently questions have arisen regarding the U.K. sampling in 2000; our results are unaffected by disregarding the 2000 scores for the United Kingdom in our analyses.

[15] France had to be excluded from the analysis because it provides no information on the school-level questionnaire. Due to their small size, Liechtenstein and Macao were also dropped.

Table 2.1
Descriptive Statistics by Country

	GDP per capita	PISA math test scores		Acad.-content autonomy		Personnel autonomy		Budget autonomy	
	2000	2000	2009	2000	2009	2000	2009	2000	2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Low-income countries[a]									
Indonesia[b]	803	366.1	371.1	.915	.695	.689	.403	.974	.809
Lower-middle-income countries[a]									
Brazil	3,701	332.8	386.0	.824	.517	.245	.156	.748	.349
Bulgaria[b]	1,600	429.6	427.9	.721	.410	.572	.796	.693	.923
Romania[b]	1,650	426.1	426.4	.737	.607	.113	.018	.996	.633
Russia	1,775	478.3	467.9	.958	.574	.704	.658	.701	.538
Thailand[b]	1,968	432.7	418.6	.961	.900	.284	.316	.896	.917
Tunisia[c]	2,033	358.9	371.5	.100	.028	.150	.016	.979	.811
Turkey[c]	4,010	423.8	445.7	.598	.218	.065	.017	.684	.773
Upper-middle-income countries[a]									
Argentina[b]	7,693	387.4	387.6	.823	.408	.212	.275	.471	.738
Chile[b]	4,877	382.9	420.7	.900	.395	.394	.635	.651	.789
Czech Republic	5,521	493.3	492.6	.878	.865	.834	.883	.991	.747
Hungary	4689	483.3	490.0	.983	.681	.705	.744	.922	.944
Latvia	3,302	461.7	481.5	.885	.413	.625	.524	.890	.826
Mexico	5,934	386.8	418.5	.661	.318	.414	.305	.773	.783
Poland[d]	4,454	470.7	494.2	.821	.750	.607	.484	.903	.264
Slovak Republic[c]	5,326	498.6	496.7	.754	.521	.798	.686	.956	.699
Uruguay[c]	6,914	421.8	427.2	.392	.216	.198	.192	.504	.577

(continued on next page)

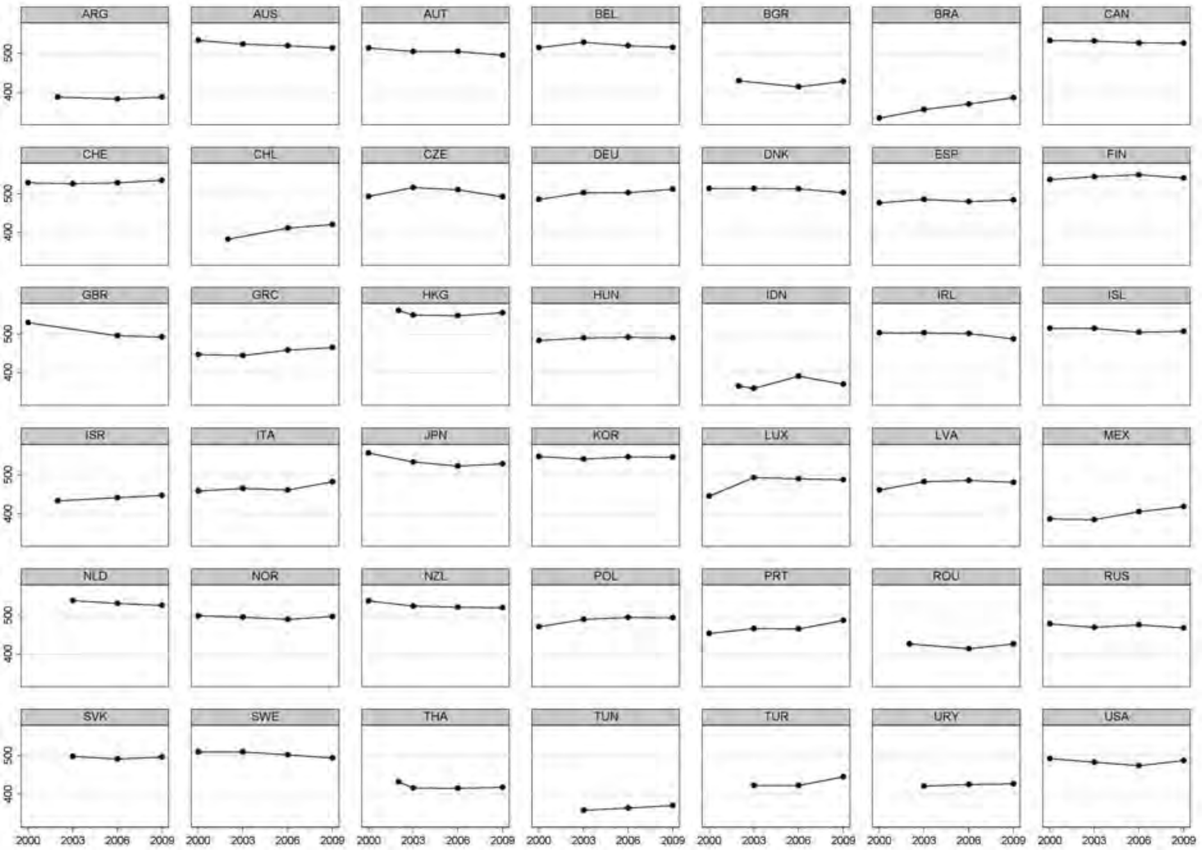
(continued)

	GDP per capita	PISA math test scores		Acad.-content autonomy		Personnel autonomy		Budget autonomy	
	2000	2000	2009	2000	2009	2000	2009	2000	2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
High-income countries[a]									
Australia	21,768	533.7	514.6	.933	.708	.389	.373	.996	.935
Austria	23,865	514.2	495.3	.700	.569	.076	.069	.925	.845
Belgium	22,665	515.2	515.7	.726	.561	.512	.381	.992	.672
Canada	23,559	533.0	526.3	.759	.323	.577	.315	.987	.763
Denmark	29,992	513.7	503.2	.889	.684	.551	.596	.979	.982
Finland	23,514	536.4	540.4	.954	.617	.181	.211	.987	.926
Germany	23,114	485.5	512.1	.552	.644	.060	.176	.956	.975
Greece	11,500	447.3	465.4	.902	.048	.689	.027	.947	.858
Hong Kong[b]	25,374	560.5	554.7	.991	.871	.586	.520	.979	.911
Iceland	30,951	515.0	507.4	.797	.675	.517	.506	.871	.775
Israel[b]	19,836	433.6	447.4	.910	.506	.740	.379	.950	.659
Ireland	25,380	503.0	487.3	.781	.687	.461	.327	1	.898
Italy	19,269	458.8	483.3	.716	.702	.057	.060	.571	.832
Japan	36,789	556.8	529.2	.988	.919	.328	.324	.912	.903
Korea	11,346	547.6	545.9	.973	.887	.234	.207	.947	.884
Luxembourg	46,457	446.1	488.2	.000	.140	.000	.172	1	.809
Netherlands[c]	24,179	538.1	525.9	.978	.922	.939	.896	.988	1
New Zealand	13,336	537.9	519.9	.957	.901	.586	.547	1	.993
Norway[c]	37,472	498.7	497.5	.571	.506	.325	.403	.982	.884
Portugal	11,443	453.4	487.3	.582	.382	.068	.107	.949	.933
Spain	14,421	476.4	483.7	.800	.538	.234	.185	.982	.959
Sweden	27,879	509.7	493.9	.879	.727	.804	.768	.993	.930
Switzerland	34,787	528.3	535.0	.381	.298	.526	.476	.869	.849
United Kingdom	25,089	529.7	492.5	.978	.871	.854	.719	.999	.946
United States	35,080	492.6	487.4	.912	.552	.867	.716	.987	.858
Country average	16,317	477.3	477.7	.780	.571	.445	.394	.902	.811

Notes: PISA data: Country means, based on non-imputed data for each variable, weighted by sampling probabilities.
[a] Country classification according to World Bank classification in 2002.
[b] PISA data refer to 2002 instead of 2000.
[c] PISA data refer to 2003 instead of 2000.
[d] Autonomy data refer to 2002 instead of 2000.

with increases surpassing one quarter of a standard deviation), while others saw substantial decreases (most notably the United Kingdom and Japan with decreases surpassing one quarter of a standard deviation).

Figure 2.1. Performance on the PISA Math Tests, 2000-2009.



Notes: Country mean performance in the PISA math test. Own depiction based on PISA tests conducted in 2000/2002, 2003, 2006, and 2009.

As with all such surveys, the dataset of all students with performance data has missing values for some background questions, although with few exceptions this is five percent or less for variables included in our analysis (see Appendix Table A2.1). Yet, since we consider a large set of explanatory variables and since a portion of these variables is missing for some students, dropping all student observations with any missing value would result in substantial sample reduction. We therefore imputed values for missing control variables by using the country-by-wave means of each. To ensure that imputed data are not driving our results, all our regressions include an indicator for each variable with missing data that equals one for imputed values and zero otherwise.

We combine the student and school data with additional country-level data. GDP per capita, measured in current US$, is provided by the World Bank and OECD national accounts data files. Data on annual expenditure per student in lower secondary education in 2000, 2003, and 2006 are taken from the OECD Education at a Glance indicators (see Organisation for Economic Co-operation and Development, 2010). Data on the existence of curriculum-based external exit exams are an updated version of the data used by Bishop (2006).

2.3.2 Measuring School Autonomy

We construct our measures of school autonomy for each country from the background questionnaires of the four PISA studies.[16] In all test waves, principals were asked to report the level of responsibility for different types of decisions regarding the management of their school. We make use of six decision-making types: 1. Deciding which courses are offered; 2. Determining course content; 3. Choosing which textbooks are used; 4. Selecting teachers for hire; 5. Establishing teachers' starting salaries; and 6. Deciding on budget allocations within the school.

In 2000 and 2003, principals were asked, "In your school, who has the main responsibility for ...". For each of the enumerated areas, principals had to tick whether decisions were mainly a responsibility of the school's governing board, the principal, department heads, or teachers as opposed to not being a responsibility of the school. Similarly, in 2006 and 2009, principals were asked who has a considerable responsibility for the enumerated tasks and had to choose whether the regional or national education authority as opposed to the principal or teachers had considerable responsibility.[17] In all four waves, respondents were explicitly allowed to tick as many options as appropriate in each area.

For each area, we begin by constructing a variable indicating full autonomy at the school level, which equals one if a school entity – the principal, the school's board, department heads, or teachers – is the only one to carry responsibility (and zero otherwise). Thus, as

[16] Measures of school autonomy could be developed from a variety of sources including the surveys of *Education at a Glance* (e.g., Organisation for Economic Co-operation and Development, 2008a) or of the European Commission (Eurydice, 2007). These sources, however, do not cover all of the countries with achievement data and do not provide the data on timing of implementation that we need. We do provide information below on how they relate to our measures.

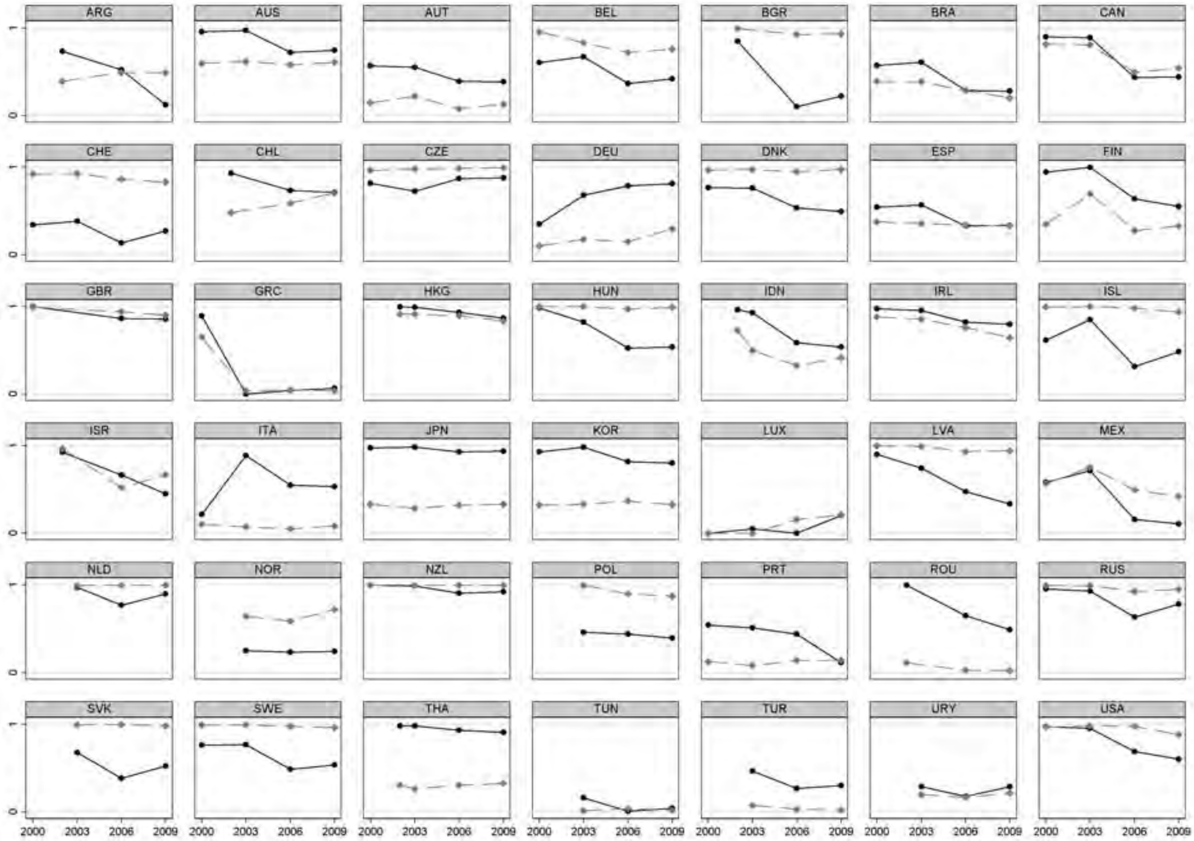
[17] See Table A2.2 in the Appendix for an overview of the answer options and a discussion on their comparability across the PISA waves.

soon as responsibility is also carried by external education authorities, we do not classify a school as autonomous. (As part of the robustness checks below, for each area we also construct a variable indicating whether the school has any influence on the decision-making process as opposed to exercising full responsibility.) Then, because our interest is focused on countries' institutional structures, we aggregate across all schools in a country to obtain the share of schools with full autonomy in each of the areas. As will be made explicit in the next section, we do not emphasize the individual school measures of autonomy in the modeling of achievement because of concerns about introducing selection bias and because of the possibility of general-equilibrium effects but we do provide the results of using the disaggregated measures of autonomy.

Figure 2.2 shows illustrative graphs across the four waves of aggregate autonomy for determining courses offered and for hiring in each country. While many countries have rather flat profiles of autonomy over time, there are also clear movements that differ between the two autonomy areas. For example, among low-achieving countries, Brazil, Chile, and Mexico have seen strong reductions in course autonomy, but smaller reductions (or even increases) in hiring autonomy. Similarly, among medium-achieving countries, Greece, Portugal, and to a lesser extent Turkey have reduced course autonomy, but this is not the case for hiring autonomy in Portugal and Turkey. At a higher level of achievement, Germany has increased school autonomy, particularly in course offerings, whereas countries such as Great Britain, Australia, Denmark, Ireland, and Sweden have all seen slight decreases in the autonomy measures.[18]

[18] While it is beyond the scope of this chapter to provide anecdotal narratives of specific policy reforms that underlie the patterns documented in the PISA data, there are many instances where main policy movements can be directly linked to the overall pattern of the PISA autonomy data. For example, based on assessments by country officials, the Organisation for Economic Co-operation and Development (2004) notes that "For example in Greece, central government had responsibility for 25% more decisions in 2003 than it did in 1998" (p. 428), quite consistent with the trend towards reduced implemented autonomy shown in the PISA data. Similarly, in Germany the increase in course-offering autonomy in the early 2000s reflects the change in governance philosophy in many German states towards "New Public Management" practices, including decentralization and introduction of school autonomy in particular in developing own course profiles (e.g., Weiß, 2004; Aktionsrat Bildung, 2010). Likewise, the increase in teacher-hiring autonomy between 2006 and 2009 likely reflects the fact that North Rhine-Westphalia enacted a new schooling law that for the first time assigned autonomy to schools in advertising open positions and hiring their own teachers (see *Schulgesetz für das Land Nordrhein-Westfalen*, Section 57, clause 7). Similarly, the decline in local decision-making about local course offerings in the U.S. is consistent with the expansion of state standards following the introduction of federal accountability legislation (No Child Left Behind) in 2002.

Figure 2.2. School Autonomy over Courses and over Hiring, 2000-2009.



Notes: Straight black lines: autonomy in deciding which courses are offered. Dashed gray lines: autonomy in selecting teachers for hire. Own depiction based on school background questionnaires in the PISA tests conducted in 2000/2002, 2003, 2006, and 2009.

Table 2.2 presents correlations among the six autonomy areas, both in their 2009 levels and in their difference between 2000 and 2009 (which provides the main source of identification in our analysis). Obviously, the three autonomy areas on decisions that are related to academic content – namely courses offered, course content, and textbooks used – are highly correlated among each other, both in levels and in changes. Also, the two autonomy areas on personnel decisions – hiring teachers and establishing their starting salaries – are strongly related. As a consequence, we combine the three variables of courses offered, course content, and textbooks used into one category of autonomy regarding academic content by using their arithmetic mean. Similarly, the mean of hiring teachers and establishing their starting salaries represents our measure of autonomy in personnel decisions.[19] Since autonomy on budget allocations is not correlated with any of the other

[19] Results are very similar if, rather than using the mean across the autonomy categories, we use the share of schools in a country that have autonomy in two or three of the subcategories of the combined variables. In the Appendix, we also report results for the six separate autonomy categories.

autonomy areas (apart from the personnel areas when considered in differences rather than levels), we retain it as a separate third autonomy category.

These autonomy measures are best thought of as the school principals' views on the reality of local decision-making, so that they should be interpreted as representing autonomy "as implemented" instead of autonomy "as prescribed." Nonetheless, it is possible to relate these to measures of autonomy from *Education at a Glance* (EAG) and from Eurydice that come from surveys of national officials. The EAG measure of autonomy in instruction is correlated between 0.52 and 0.65 with our measure of autonomy of academic content across the three comparison years that are available.[20] The correlations between the measures of personnel autonomy range from 0.35 to 0.64. The PISA budget autonomy measure is only weakly correlated with the EAG measures, possibly because EAG does not directly cover budget autonomy. For its part, Eurydice (2007), in its rich discussion of different structures and movements toward local autonomy, makes it clear that there is a substantial difference between legislation that allows or requires more autonomous decisions and the actual adoption of local decision-making. In particular, from the descriptions a variety of laws that called for greater local decision-making and did not emanate from the localities themselves, it was unclear exactly when and how far any implementation went. The lack of information on the pattern of implementation plus the general perspective of Eurydice (2007) on the broader trends as opposed to the degree of autonomy at any specific times makes it impossible to correlate their data with our measures of autonomy.

2.3.3 Descriptive Statistics

Table 2.1 presents country-level means of the three autonomy measures, as well as mean PISA math scores, in 2000 and 2009. Throughout the chapter, our analysis focuses on mathematical literacy, which is generally viewed as being most readily comparable across countries; however, we also report main results in reading and science. Table A2.1 in the Appendix reports pooled international descriptive statistics for all variables employed in the analysis.

[20] In these, we correlate EAG in 1998 with PISA in 2000 (21 countries); EAG in 2003 with PISA in 2003 (23 countries); and EAG in 2007 with PISA in 2006 (21 countries). The highest correlation (0.65) occurs in 2003 when both are measured at the same time (see Table A2.3).

Table 2.2
Country-Level Correlation Matrix of Autonomy Measures

(A) 2009 levels	Courses	Content	Textbooks	Hiring	Salaries	Budget	Academic-content	Personnel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School autonomy over courses	1							
School autonomy over content	0.739***	1						
School autonomy over textbooks	0.511***	0.598***	1					
School autonomy over hiring	0.385**	0.384**	0.366**	1				
School autonomy over salaries	0.417***	0.398***	0.209	0.576***	1			
School autonomy over budget allocations	0.274*	0.060	0.228	0.089	0.186	1		
Academic-content autonomy	0.865***	0.905***	0.817***	0.438***	0.395***	0.215	1	
Personnel autonomy	0.445***	0.436***	0.340**	0.933***	0.832***	0.143	0.472***	1

(B) 2000–2009 differences	Courses	Content	Textbooks	Hiring	Salaries	Budget	Academic-content	Personnel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School autonomy over courses	1							
School autonomy over content	0.461***	1						
School autonomy over textbooks	0.560***	0.547***	1					
School autonomy over hiring	0.559***	0.342**	0.688***	1				
School autonomy over salaries	0.316*	0.295*	0.730***	0.749***	1			
School autonomy over budget allocations	0.066	-0.150	0.199	0.403**	0.427**	1		
Academic-content autonomy	0.846***	0.813***	0.811***	0.626***	0.503***	0.030	1	
Personnel autonomy	0.454***	0.338**	0.760***	0.921***	0.948***	0.445***	0.597***	1

Notes: Correlation coefficient of country-level autonomy measures across 42 countries. Data for Argentina, Bulgaria, Chile, Hong Kong, Indonesia, Israel, Romania, and Thailand refer to 2002 instead of 2000. Data for Slovak Republic, Tunisia, Turkey, and Uruguay refer to 2003 instead of 2000. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

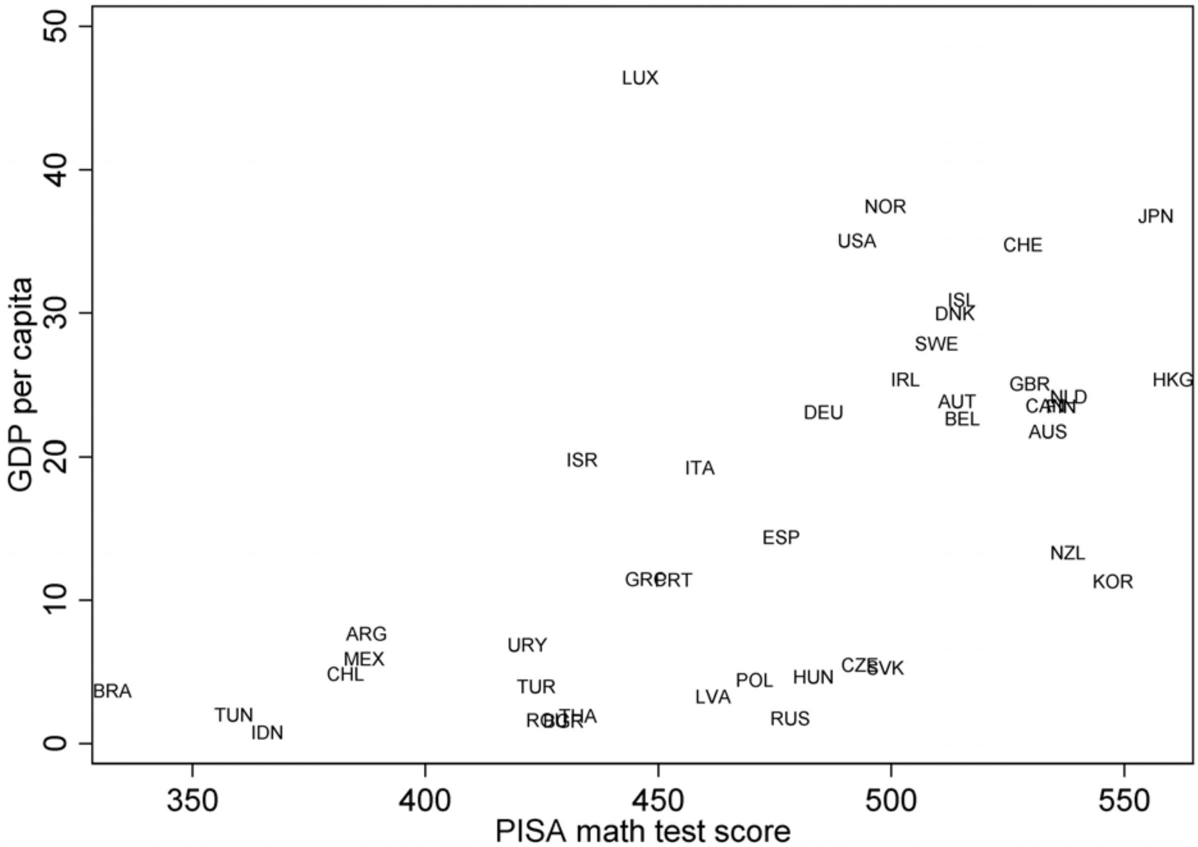
Table 2.1 also shows a country's GDP per capita in 2000, our main measure of initial level of development. Figure 2.3 plots this measure of initial economic development against initial educational achievement, measured as the PISA math score in 2000. There is a strong relation between the initial levels of economic and educational development, which we will further explore below. Most importantly, the figure visualizes where different countries stand on these measures of initial development, which is informative for our analysis of heterogeneity across initial country situations below.

From Figure 2.1, we can assess the development of PISA math test scores across waves for all 42 countries. Among the low-performing countries with initial test scores below 400 points, Brazil, Chile, Mexico, and moderately Tunisia managed to increase their test scores over time, whereas Argentina's and Indonesia's achievement is mostly flat. Within the group of medium performers, Greece, Italy, Israel, Portugal, and Turkey show a slightly positive trend, whereas Thailand followed a slight downward trend. Among the countries with initially relatively high scores, only Germany shows a consistent upward trend, whereas Great Britain and Japan, and to a lesser extent Australia, Austria, Denmark, Ireland, New Zealand, and Sweden, show a downward trend. The other countries are mostly flat.[21]

Comparing these achievement trends to the autonomy trends seen in Figure 2.2, there are many examples where the combined achievement and autonomy trends are consistent with increased autonomy, particularly over academic content, being bad in low-performing but good in high-performing countries. For example, starting at a low level of achievement, the increasing achievement levels of Brazil, Chile, and Mexico are accompanied by reductions in autonomy of their schools in particular over course offerings. Similarly, Greece, Portugal, and Turkey have reduced their course autonomy and slightly increased their achievement. By contrast, Thailand – which had quite flat autonomy – saw mostly flat achievement. Finally, at a higher level of initial achievement, Germany's increased autonomy, particularly over course offerings, goes along with consistent increases in achievement. Great Britain, Australia, Denmark, Ireland, and Sweden all slightly reduced their autonomy, which is mirrored by slightly decreasing achievement.

[21] These trends on just the PISA tests for 2000-2009 are very consistent with the longer trends from 1995-2009 that also include scores on the other international assessments of TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study), see Hanushek et al. (2012).

Figure 2.3. Development Level and PISA Performance, 2000.



Notes: Test scores for Argentina, Bulgaria, Chile, Hong Kong, Indonesia, Israel, Romania, and Thailand refer to 2002. Test scores for Slovak Republic, Tunisia, Turkey, and Uruguay refer to 2003.

2.4 Empirical Model

To test the effect of autonomy on student achievement and its dependence on a country's development level more formally, we make use of the education production function framework introduced above. The empirical issues can be most easily seen from a simple linear formulation which now introduces a time dimension to the analysis:

$$T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \epsilon_{cti} \tag{2.3}$$

where achievement T in country c at time t for student i is a function of a country's institutions I (here autonomy), the inputs from a student's family (F) and from schools (S), and an error term, ϵ_{cti}. We start our exposition with a linearized and additive version of the model, but our analyses below will test for rich multiplicative interactions of the institutional effect with other input factors. Our interest is estimating $\alpha = \partial T/\partial I$, the impact of local autonomy on achievement holding constant other inputs. For this, we have the panel data from PISA that has individual-level data about T, F, and S and data about institutions I aggregated at the country level.

Our approach to identify the impact of institutions is best seen by expanding the error term:

$$\epsilon_{cti} = \eta_c + \eta_{ct} + \eta_{cti} \tag{2.4}$$

where η_c is a time-invariant set of cultural and educational factors for country c (such as awareness of the importance of education, the commitment of families to their children's education, or more generally the state of development of societal and economic institutions); η_{ct} is a time-varying set of aggregate educational factors for country c (such as changes in spending levels or private involvement); and η_{cti} is an individual-specific, time-varying error.

The key to identification of α, the parameter of interest, is that ϵ_{cti} is orthogonal to the included explanatory factors and, importantly, to the measure of local autonomy. The formulation in Equation 2.4 shows the main elements of our approach. First, at the individual student and school level, there are concerns about selection bias, reflecting unmeasured attributes of schools or students in circumstances with varying local decision-making.[22] If, for example, particularly good students are attracted to schools with more local autonomy, η_{cti} would tend to be correlated with I, leading to bias in the estimation of α. But, by aggregating over all schools in the country and measuring autonomy by the proportion of schools with local autonomy, we eliminate the selection bias from school choice. The aggregation also allows us to capture any general-equilibrium effect whereby, for example, autonomy of one school may elicit competitive responses from schools that do not have autonomy themselves.

Second, with the panel data, we can include country fixed effects, μ_c, which effectively eliminate any stable country-specific factors contained in η_c,[23]

$$T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \mu_c + \mu_t + \nu_{cti} \tag{2.5}$$

[22] These concerns are central to the interpretation of most within-country analyses of decentralization. Some micro-evaluations do, however, circumvent these problems by focusing on external policy changes; e.g., Galiani et al. (2008).

[23] The estimation also includes time fixed effects to allow for any common shocks across waves.

By implication, the estimation of α is based upon variations in autonomy over time, since time-invariant institutional features are absorbed into the country fixed effect. The relevant variation with which we estimate α is within-country changes for our sample of PISA countries.

The most significant remaining issue is whether there are time-varying country factors (η_{ct}) that are correlated with the pattern of local autonomy in the country. The underlying identifying assumption is that there are no educationally important time-varying country factors that are correlated with variation in the institutional input, I. We will partially test this by including several additional time-varying factors of countries' education systems, C_{ct}, in the analysis:

$$T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \beta_C C_{ct} + \mu_c + \mu_t + \nu_{cti} \tag{2.6}$$

While there are of course a variety of factors that could enter, our approach is to use our rich survey dataset to eliminate the most significant characteristics of the schools and the parental population.

Other details are also important. In order to obtain the best estimates of α, we attempt to eliminate as much other variation in test scores as possible by estimating the β parameters for family and school effects on a large set of individual measures and by conducting the estimation at the individual student level. Additionally, the limited variation in institutional factors – which occurs at the country level – means that it is hard to simultaneously estimate measures of alternative forms of local decision-making. As a result, most of our analysis sequentially estimates models with combined autonomy measures, although we also report specifications that include several autonomy measures together.

A central component of the analysis, motivated by the conceptual model and by the prior within-country analyses, is the possibility of significant interactions of institutional factors with other institutions or country-specific elements such as school accountability systems or level of capacity and stage of development. We pursue this parametrically by interacting I, the specific measure of autonomy in each model, with the initial level of development (of the country and/or educational system), D_c:

$$T_{cti} = \alpha_1 I_{ct} + \alpha_2 (I_{ct} \times D_c) + \beta_F F_{cti} + \beta_S S_{cti} + \beta_C C_{ct} + \mu_c + \mu_t + \nu_{cti} \tag{2.7}$$

In this model, which represents our main specification, the effect of autonomy reforms is allowed to differ depending on the surrounding conditions captured by D_c. We can then test our main conceptual proposition that autonomy is beneficial for student achievement in otherwise well-functioning systems but detrimental in dysfunctional systems.

2.5 Results

2.5.1 Main Results

Conventional estimation identifies the effect of autonomy from the cross-sectional variation. For comparison to our identification below, such models are reported in Table 2.3. A simple pooled cross-section with school autonomy measured at the individual level shows a positive association of the three areas of autonomy with student achievement in math (significant for academic-content and budget autonomy), after controlling for standard measures of family and school background (column 1). There is little indication that this association differs across levels of development, although the positive association of academic-content autonomy seems to increase slightly with a country's development level, measured by the initial GDP per capita in 2000 (column 2). The average cross-sectional associations vanish when country fixed effects are added to the model (column 3). At least for academic-content autonomy, there is a significant positive interaction between initial GDP per capita and autonomy in the model with country fixed effects (column 4). However, in models with country-by-year fixed effects (column 5) that effectively look just at within-country variation in individual school autonomy and that eliminate any time influences on the estimates, there is no indication of any influences of local decision-making on student achievement. The main concern with these estimates is that they are heavily influenced by potential selection biases arising from the specific schools that indicate having local autonomy.

When we avoid these within-country selection problems by averaging the autonomy measures at the country level (while keeping all other variables at the individual level), the estimates of the impact of autonomy increase substantially (column 6). Again, there

Table 2.3

Conventional Cross-Sectional Estimation of the Effect of School Autonomy on Student Achievement

Autonomy measured at level:	School		School			Country		Country
Country fixed effects:	No		Yes			No		Yes
Country-by-year fixed effects:	No		No	No	Yes	No		No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Academic-content autonomy	20.713***	13.539*	-1.387	-4.792**	0.269	47.201***	37.114**	-20.556
	(6.181)	(7.455)	(2.106)	(2.171)	(2.459)	(11.257)	(14.076)	(12.627)
Academic-content autonomy x Initial GDP p.c.		0.771*		0.438***	0.009		0.908	
		(0.455)		(0.137)	(0.193)		(0.616)	
R^2	0.312	0.315	0.384	0.384	0.392	0.319	0.321	0.384
Personnel autonomy	9.640	10.479	0.844	2.383	3.298	24.701*	24.913*	-0.180
	(7.015)	(7.586)	(3.483)	(4.983)	(5.182)	(13.492)	(13.313)	(11.708)
Personnel autonomy x Initial GDP p.c.		-0.103		-0.199	-0.343		-0.024	
		(0.535)		(0.404)	(0.405)		(1.055)	
R^2	0.310	0.310	0.384	0.384	0.392	0.312	0.312	0.384
Budget autonomy	7.549*	5.411	2.350	2.366	3.657*	32.987	31.239	-7.163
	(4.248)	(4.694)	(1.536)	(1.771)	(1.869)	(25.976)	(25.228)	(10.162)
Budget autonomy x Initial GDP p.c.		0.493		-0.003	-0.115		1.127*	
		(0.336)		(0.132)	(0.132)		(0.631)	
R^2	0.310	0.310	0.384	0.384	0.392	0.311	0.313	0.384

Notes: Each column-by-panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability. In columns 2 and 4, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 2.4
Panel Fixed-Effects Results on the Effect of School Autonomy on Student Achievement by Development Level

	Estimation result	Details on autonomy effect at different levels of GDP per capita		
		GDP p.c. at which autonomy effect switches sign	Effect in country with minimum GDP p.c.	Effect in country with maximum GDP p.c
	(1)	(2)	(3)	(4)
Academic-content autonomy	-34.018***	19,555	-55.205***	52.754***
	(12.211)		(14.471)	(16.670)
Academic-content autonomy x Initial GDP p.c.	2.944***			
	(0.590)			
R^2	0.385			
Personnel autonomy	-17.968	13,413	-41.854**	79.861***
	(14.071)		(19.201)	(28.647)
Personnel autonomy x Initial GDP p.c.	3.319***			
	(1.106)			
R^2	0.384			
Budget autonomy	-6.347	11,449	-19.576	47.833**
	(9.363)		(13.282)	(20.221)
Budget autonomy x Initial GDP p.c.	1.838**			
	(0.796)			
R^2	0.384			

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. "Maximum GDP p.c." refers to Norway. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Complete model of the first specification displayed in Table A2.1. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

is little sign of effect heterogeneity across development levels (column 7). However, results change dramatically when, consistent with our identification strategy, we focus on within-country changes over time with autonomy aggregated to the national level. The cross-sectional associations vanish, with point estimates turning negative, once country fixed effects are added (column 8), where the autonomy effect is now identified from aggregate within-country variation over time. Still, this average effect may hide substantial heterogeneity of the autonomy effect across countries.

Thus, Table 2.4, which shows our main results, adds an interaction term of autonomy with initial GDP per capita to the panel specification with country fixed effects and with autonomy measured at the country level.[24] The results indicate clear evidence of substantial effect heterogeneity for all three areas of autonomy: The autonomy effects become significantly more positive with increasing initial GDP per capita. GDP per capita is centered at $8,000 (in 2000) in this specification, implying that the main effect reflects the impact of autonomy on student achievement in a country at the upper end of the upper-middle-income category of countries such as Argentina (see Table 2.1 and Figure 2.3).[25]

As indicated by the negative main effect, a country near Argentina's level of development that increased its academic-content autonomy over time would expect to see a significant and substantial drop in achievement. In such a country, going from no autonomy to full autonomy over academic content would reduce math achievement by 0.34 standard deviations according to this model. Moreover, the significant positive interaction indicates that the autonomy effect is significantly negative for all low- and middle-income countries in our sample. At the extreme of the poorest country in our sample (Indonesia at $803 GDP per capita in 2000), the negative effect of academic-content autonomy reaches 0.55 standard deviations (column 3).

By contrast, the effect of academic-content autonomy turns significantly positive in most of the high-income countries. Near the top of the income distribution by countries

---

[24] Table A2.1 in the Appendix shows the coefficients of the control variables in this specification for the academic-content autonomy category.

[25] This possibility of differential impacts depending on decision-making capacity was originally suggested by micro-evaluation studies (see Galiani and Perez-Truglia, 2011), but the cross-country results here do not just reflect variations in outcomes that arise from differential impacts by socio-economic status within countries. We find that measures of variations in family backgrounds within countries never enter significantly into our models and do not affect our main results.

(Norway at $37,472 GDP per capita in 2000), the positive effect of academic-content autonomy is as large as 0.53 standard deviations (column 4).[26] The level of 2000 GDP per capita at which the autonomy effect switches its sign from negative to positive is $19,555 (column 2). As is evident from Table A2.4 in the Appendix, this pattern holds separately for all three categories of autonomy – course offerings, course content, and textbooks – contained in the aggregated measure of academic-content autonomy in this table.

As the lower two panels show, the basic pattern of results is quite similar in the other two areas of autonomy – personnel and budget autonomy. The autonomy effect increases significantly with initial GDP per capita, and there is a large and significant positive autonomy effect for rich countries. The only difference from the academic-content autonomy category is that the negative effect in the categories of personnel and budget autonomy is smaller and not statistically distinguishable from zero at the upper end of the upper-middle-income countries. For budget autonomy, the negative autonomy effect does not reach statistical significance for even the poorest country in our sample.

The substantial correlation between the different categories of autonomy limits the extent to which we can distinguish among the three categories, but Table 2.5 presents models with pairs of two autonomy variables, as well as all three of them, combined. When academic-content autonomy is included together with the other autonomy categories, only the interaction of academic-content autonomy with initial GDP per capita retains statistical significance. When only personnel and budget autonomy are included, the interaction of initial GDP per capita with personnel autonomy is statistically significant but the interaction with budget autonomy is not. Given the high correlation of academic-content and personnel autonomy (see Table 2.2) and the size of the standard errors, multicollinearity does not allow us to rule out a substantial positive interaction for personnel autonomy. However, given that the correlation of budget autonomy with the other autonomy categories is quite low, these specifications tentatively indicate that budget autonomy has no separate effect once the other autonomy categories are considered.[27] Therefore, in the remainder of the chapter, we focus on the two aggregated measures of autonomy over academic content and over personnel.

---

[26] We exclude Luxembourg from these calculations because of its size and concerns about the measurement of its income. If we evaluated the impacts at Luxembourg levels, the estimated effects would be considerably larger (see Table 2.1).

[27] The significant correlation between the *change* in budget and personnel autonomy (panel B of Table 2.2) suggests that there is still some possibility that multicollinearity is driving the lack of significance.

Table 2.5
Robustness: Impact of Including Several Autonomy Measures Together in the Same Estimation

	(1)	(2)	(3)	(4)
Academic-content autonomy	-42.013***	-41.012***	-33.732***	
	(14.248)	(14.120)	(12.054)	
Academic-content autonomy x Initial GDP p.c.	2.658***	2.736***	2.888***	
	(0.674)	(0.676)	(0.616)	
Personnel autonomy	17.830	13.897		-14.998
	(16.250)	(14.449)		(13.534)
Personnel autonomy x Initial GDP p.c.	0.212	0.333		2.868**
	(1.293)	(1.256)		(1.184)
Budget autonomy	-5.370		-1.918	-2.525
	(8.858)		(8.106)	(9.205)
Budget autonomy x Initial GDP p.c.	0.240		0.151	1.049
	(0.898)		(0.832)	(0.923)
R^2	0.385	0.385	0.385	0.385

Notes: Each column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

2.5.2 Robustness Tests

Several extended tests confirm the robustness of our main finding. The various modifications for measurement issues and estimation samples leave our basic findings intact.

The first set of robustness tests relates to the measurement of variables. The main results prove quite independent of the specific way in which the interaction with initial GDP per capita is specified. As shown in the first three columns of Table 2.6, the basic result does not change when initial GDP per capita is not measured linearly, but instead in logs; as a dummy for countries with a GDP per capita higher than $8,000 (roughly the upper end of the upper-middle-income category of countries in our sample); or as a dummy for countries with higher-than-median GDP per capita in our sample (which is at $14,000).

Note that the reported specifications control for a country's current GDP per capita. Adding the change in per capita GDP or its growth rate has no substantive effect on the estimates. Neither does leaving current GDP per capita out of the model altogether (because this control might confound the effects of autonomy reforms) change the substantive results (see Table A2.5 in the Appendix).

Our main model includes measures of school characteristics, but the final columns of Table 2.6 show that results are robust to alternative treatments of school controls. First, giving autonomy to schools may mean that schools use their autonomy to alter other school characteristics, such as reducing the school size or raising teacher education requirements. Such changes would thus be channels through which school autonomy affects student outcomes. In this perspective, these school measures should not be controlled for in the estimation. As is evident in column (4), leaving the school-level variables out of our basic model does not affect our qualitative results.

Second, there may be a concern that other school reforms may have coincided with the autonomy reforms that identify our main result. To capture such other reforms, column (5) includes all school variables measured as country averages, aggregating them to the same level at which the autonomy variables are measured. Despite concerns with statistical power with a large number of country-level variables, the qualitative results for autonomy again remain the same.

Table 2.6

Robustness: Different Forms of Measuring Initial GDP per Capita and Different School Controls

Measure of initial GDP per capita:	log GDP p.c.	Dummy for GDP p.c. above \$8,000	Dummy for GDP p.c. above median (\$14,000)	GDP per capita	
School controls measured:		at school level		no school controls	as country means
	(1)	(2)	(3)	(4)	(5)
Academic-content autonomy	-74.059***	-58.521***	-37.826**	-29.920***	-30.572**
	(21.937)	(14.222)	(17.193)	(10.443)	(11.461)
Academic-content autonomy x Initial GDP p.c.	24.071***	60.362***	60.865***	2.646***	2.200***
	(7.466)	(13.013)	(12.093)	(0.539)	(0.731)
R^2	0.385	0.385	0.385	0.373	0.374
Personnel autonomy	-55.008*	-41.921*	-21.154	-15.813	-15.642
	(29.136)	(24.595)	(13.863)	(15.393)	(13.609)
Personnel autonomy x Initial GDP per capita	23.004**	62.247**	64.163***	2.750***	1.972
	(10.172)	(27.308)	(20.026)	(0.968)	(1.315)
R^2	0.384	0.384	0.384	0.372	0.373

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In columns 4 and 5, initial GDP per capita is centered at \$8,000 (measured in \$1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Autonomy reforms might also have coincided with expenditure reforms across countries. Because there is no consistent data on expenditure per student for all countries and waves, our basic model does not control for expenditure per student. But for the waves 2000-2006, we have consistent data on annual expenditures per student in lower secondary education for a subset of (mostly OECD) countries. The first column of Table 2.7 shows that our basic results hold similarly in this subset of country-by-wave observations. Column (2) adds the expenditure variable to this model, and the qualitative results are unaffected. Changes in expenditure per student are actually significantly negatively related to changes in student achievement, which dilutes concerns about the lack of expenditure controls in our basic specification. The coefficient on expenditures may capture forces that push for increased spending but that at the same time lower the efficiency of their use.[28]

The other four columns of Table 2.7 test for robustness in different sub-samples. The PISA math test was scaled to have mean 500 and standard deviation 100 across the OECD countries in 2000 and in 2003 each, and it was designed psychometrically to have a common scale since 2003. Column (3) shows that results are qualitatively unaffected when dropping the 2000 wave and restricting the analysis to the three waves since 2003 in which the tests are psychometrically scaled to be inter-temporally comparable.

In order to ensure that the effect is identified only from long-term changes and not driven by short-term oscillations, column (4) restricts the analysis to waves 2000 and 2009. When identified from the nine-year differences in autonomy and test scores, results are even more pronounced than in the four-wave specification.

Our main specification employs an unbalanced sample, as some countries did not participate in all four PISA waves (see Figure 2.1). Column (5) of Table 2.7 replicates our analysis for the fully balanced sample of 29 countries with achievement and autonomy data in all four PISA waves. Again, qualitative results are the same. Column (6) restricts the sample to OECD countries, without substantive changes in results.

Additional robustness tests show that results also do not hinge on any specific country being included in the estimation. All results are robust when we drop one country at a time from the estimation sample.[29] In particular, results look very similar when Luxembourg –

[28] As reviewed in Hanushek and Woessmann (2011b), international comparative studies of the impact of expenditures provide mixed results but tend to indicate no consistent relationships between spending and international test scores.

[29] Detailed results are available from the authors on request.

Table 2.7

Robustness: Including Expenditure per Student and Different Subsamples of Waves and Countries

Sample:	Sample with expenditure data		Waves 2003, 2006, and 2009	Waves 2000 and 2009	Balanced panel	OECD countries
	(1)	(2)	(3)	(4)	(5)	(6)
Academic-content autonomy	-31.849	-24.753	-32.263***	-54.262**	-36.980**	-28.218**
	(22.327)	(17.526)	(10.661)	(22.019)	(14.97)	(13.324)
Academic-content autonomy *x* Initial GDP p.c.	2.976**	2.645***	1.948***	4.050***	2.958***	2.529***
	(1.106)	(0.913)	(0.495)	(1.132)	(0.702)	(0.760)
Expenditure per student (in $1,000)		-11.375**				
		(4.826)				
R^2	0.362	0.363	0.389	0.382	0.362	0.308
Personnel autonomy	-52.044***	-39.557**	-28.282	-7.312	-45.458***	-21.601
	(14.381)	(15.577)	(20.462)	(18.926)	(15.033)	(13.879)
Personnel autonomy *x* Initial GDP p.c.	2.973***	1.977*	3.006**	3.441*	4.442***	3.060**
	(1.002)	(0.977)	(1.204)	(1.914)	(1.267)	(1.498)
Expenditure per student (in $1,000)		-11.867*				
		(5.932)				
R^2	0.361	0.362	0.389	0.379	0.361	0.308
Students	392,862	392,862	931,831	435,502	846,221	835,478
Countries	25	25	42	36	29	31
Countries-by-waves	67	67	120	72	116	116

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in the sample indicated on top of each column. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

a slight outlier with the highest GDP per capita (see Figure 2.3) – is excluded from the sample.

Our main results consider achievement of students in all types of schools (public and private) in order to capture any general equilibrium effects of local decision-making. However, because the autonomy reforms generally considered apply just to the public schools, we also estimate the models just for the students in public schools in each country. The pattern and significance of results remains unchanged from our preferred estimates in Table 2.4 (see Table A2.6 in the Appendix).

Finally, results are also very similar when we separate the student and country estimations into two steps. In the two-step model, test scores are "cleared" from impacts of the student- and school-level controls in a first, student-level regression. The residuals of this regression, which capture that part of the test-score variation that cannot be attributed to the controls, are then collapsed to the country-by-wave level. In a second, country-level regression, we use the country-level data to run a "classical" panel fixed-effects model, where the level of observation coincides both with the level of the fixed effects and with the level at which the variables of interest are measured. Results (shown in Table A2.7 in the Appendix) are qualitatively the same as in our preferred one-step specification, and they do not differ depending on whether the model does or does not already include country fixed effects in the estimation of the first step.

2.5.3 Specification Tests

Our identification derives from country-level variation in autonomy over time and its interaction with initial development levels in a panel model with country fixed effects. To analyze the validity of the specification, we present a set of specification tests that address several possible remaining concerns with the identification and that also indicate possible channels and sources of heterogeneity in the impacts. Given that the tests corroborate our main specification mostly by producing the result of insignificant alternative effects, we simply summarize the findings here. Detailed results are available from the authors upon request.

First, our estimates combine countries across a wide range on income and development. Because the student assessments consider only students currently enrolled in school at age 15, low enrollment rates in poor countries could artificially increase test scores (presuming that the lowest achievers are the ones dropping out of school). Nonetheless, estimating our main model with a measure of school enrollment rates (taken from the PISA documentation) has virtually no impact on our estimates (see Table A2.8 in the Appendix).

A second possible concern with identification from panel variation is that variation in autonomy over time may be endogenous to the initial level of student achievement. For example, poor initial achievement might theoretically induce governments to implement decentralization – or centralization – reforms. In order to test for the empirical relevance of this concern, we estimate several models where the changes in autonomy that identify our results are regressed on initial PISA scores. Thus, we test whether the PISA score in 2000 predict the change in autonomy from 2000 to 2003 or from 2000 to 2009. We also test whether the PISA level in one cycle predicts the change in autonomy from this to the subsequent cycle in a panel model of the four PISA waves. In all tests, lagged PISA scores do not significantly predict subsequent changes in autonomy, corroborating the identifying assumption of our panel model. Similarly, initial GDP per capita was uncorrelated with changes in autonomy between 2000 and either 2003 or 2009 (see Table A2.9). Thus, neither development level nor added resources systematically relate to the patterns of change in local autonomy.[30]

A third possible concern is that the development level may interact not only with autonomy reforms, but also with other education policy measures. In other words, the heterogeneity of impact may not be specific to the dimension of school autonomy, as other policies may also be more effective within a well-functioning surrounding. To investigate this, we included in the regression interactions of initial GDP per capita with country-level measures of several other features of the school system: competition (proxied by the share of privately operated schools), funding sources (share of public funding in the school budget), school size (number of students per school), teacher education (share of certified teachers), and shortage of math teachers. Our results show that none of these variables interacts significantly with initial GDP per capita in determining student achievement, and the autonomy results remain robust when these additional interactions are included in the model.

[30] This lack of systematic relationship with country income levels can also be seen from the Eurydice (2007) descriptions of the use of local decision-making across its sample of countries.

Fourth, to investigate whether the heterogeneity of the autonomy effect is specific to the development level and does not capture heterogeneity with respect to other country characteristics, we also estimated specifications that interact autonomy with a number of other country measures. (For interactions specifically with the overall performance of the education system and with accountability, see the next section). Some of these measures may also be interpreted as possible channels through which the level of economic development may matter for the impact of autonomy on student achievement. Specifically, autonomy may interact with the size of a country, as school autonomy may mean different things in small and large countries; with its ethnic homogeneity, as autonomy may work better in homogenous societies; with a country's political regime, corruption level, or governance effectiveness, which may determine restraints on how well autonomy can work; or with a country's culture, which may be more or less complementary to autonomous decision-making. In addition, parental human capital may moderate the quality of local monitoring, their ability to pay for private schooling may affect the incentives of autonomous schools, and autonomous schools may use specific local policies.

Thus, we estimated specifications that interact autonomy with population size; with the Alesina et al. (2003) measure of ethnic fractionalization; with the Polity IV index that measures governing authority on a scale from institutionalized autocracies to consolidated democracy; with the corruption perceptions index of Transparency International; with the Governance Effectiveness Index of the World Bank's Worldwide Governance Indicators project, which aims to capture the perceived quality of public services and of policy formulation and implementation; and with the six Hofstede dimensions of national culture, in particular the measures of individualism versus collectivism (integration into groups) and of power distance (acceptance of power inequality). We also interacted autonomy with average measures of parents' human capital available in the PISA dataset (white collar occupations and books at home), with the share of private funding in the school budget, and with such school aspects as the share of certified teachers, shortages of math teachers, school size, and share of private schools.

In models that enter these interactions separately and do not include the interaction of autonomy with initial GDP per capita, there is an indication that autonomy interacts positively with democracy, government effectiveness, individualism, the share of privately operated schools, and the share of certified teachers, and negatively with population size, corruption, and acceptance of power inequality. However, in all these cases, the significance of the interaction vanishes once the interaction of academic-content autonomy with initial GDP per capita is also entered, and the latter retains statistical significance throughout.[31] Thus, while the interaction with the development level clearly entails dimensions of democracy, governance effectiveness, cultural values, and effective school environments, the overall measure of economic development in terms of GDP per capita dominates these other separate interactions. Variations in these other measures that are not correlated with the standard measure of economic development do not interact significantly with the autonomy effect.

Fifth, we test whether the autonomy effect is heterogeneous for students with different individual social backgrounds. Such heterogeneity may reflect another channel of the autonomy effect, as decentralization may work better with sophisticated parents (Galiani et al., 2008). It also provides evidence on the effect of autonomy on inequality, as differential impacts by social background would narrow or widen the performance gap between well-off and disadvantaged families. To test this, we add interaction terms between autonomy and family background measures as well as the triple interaction between autonomy, initial GDP per capita, and the family measures to our basic specification. Our measures of individual family background include parental white collar occupation, parental university education, books at home, and immigration background. For all four measures, neither the interaction with the autonomy variable nor the triple interaction is statistically significant, and point estimates suggest different directions of effects. Consequently, autonomy reforms do not seem to affect children from different background differently and thus do not seem to magnify or lessen inequality, either in developed or in developing countries.[32]

[31] Results for personnel and budget autonomy are similar, but sometimes less strong. While the negative interaction of autonomy with ethnic fractionalization is insignificant in the separate model, it turns marginally significant in the model that also includes the interaction of autonomy with initial GDP per capita (which is fully robust), indicating that autonomy may work better in ethnically more homogeneous countries.

[32] We also estimated a specification that adds an interaction of autonomy with the initial Gini coefficient of income inequality, provided by the World Bank. While the interaction of autonomy with the initial per-capita GDP level remains qualitatively unaffected, there is also some indication that academic-content autonomy is more beneficial in more equal societies. However, this pattern is not confirmed by distributional measures of family background taken from the PISA dataset that directly relate to the parents of the tested students.

2.5.4 Further Results

While the results so far relate to math achievement, which is most readily tested comparably across countries, PISA also tested students in reading and science. As shown in column (1) of Table 2.8, results are qualitatively the same in reading. This is particularly interesting because reading scores have been psychometrically scaled to be comparable over all four PISA waves. Results on academic-content autonomy are also found for science achievement, where results on personnel autonomy are less pronounced and lose statistical significance (column 2).

In our analysis so far, we have defined autonomy as a school entity having the sole responsibility for a task. Alternatively, one can consider cases where a school entity has considerable responsibility, but an authority beyond the school has considerable responsibility as well–something that one might term "joint decision-making." Conceptually, one might expect that both the negative and the positive aspects of autonomy discussed in our conceptual framework might be somewhat limited when an external authority has a joint say on a matter. To test this, we use as an alternative autonomy measure the share of schools in a country that have considerable responsibility on a task but where an external authority may also have a say.

Column (3) of Table 2.8 shows that results are considerably weaker for this "joint authority" measure of school autonomy than for the measure of "full" school autonomy used throughout this chapter. Both negative and positive effects of autonomy are reduced when external education authorities may also have a say in decision-making. Thus, the main effects of autonomy derive from independent decision-making at the school level.

Another aspect of the specific type of autonomy is the difference between legislation and implementation. As discussed in Section 2.3.2 above, the PISA-based measures of implemented academic-content and personnel autonomy show substantial correlations with the respective EAG-based measures of legislated autonomy (see Table A2.3). Although the EAG measures are available only for a limited sample of countries and years (up to 22 countries in 1998, 2003, and 2007, for a total of 57 country-by-wave observations) and their years of observation do not match the PISA observations properly, we can also estimate our panel regression models using the EAG measures as alternative autonomy measures. For the combined EAG measure of autonomy and for its domain of instruction, results

Table 2.8
Further Results: Other Subjects and Joint Authority

Subject:	Reading	Science	Math
Measurement of autonomy:	Full autonomy		Joint authority
	(1)	(2)	(3)
Academic-content autonomy	-12.938	-28.529**	-26.070*
	(8.928)	(11.484)	(13.152)
Academic-content autonomy *x* Initial GDP p.c.	2.094***	1.115**	1.185*
	(0.557)	(0.505)	(0.627)
R^2	0.351	0.337	0.384
Personnel autonomy	-6.929	-12.430	0.709
	(14.018)	(10.810)	(13.838)
Personnel autonomy *x* Initial GDP p.c.	3.098***	0.550	1.335
	(1.066)	(0.853)	(0.921)
R^2	0.351	0.336	0.384
Students	1,125,794	1,042,791	1,042,995
Countries	42	42	42
Countries-by-waves	154	155	155

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA test score in respective subject. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at \$8,000 (measured in \$1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

confirm our pattern of a significant negative autonomy effect in developing countries and a significant positive interaction with the level of development leading to a significant positive effect in developed countries (see Table A2.10 in the Appendix), despite the limited extent and match of the data. Thus, while our main results have the advantage of capturing the effect of autonomy as actually implemented, they also appear to hold when using measures of legislated autonomy, something that is more directly amenable to policymakers.

2.6 Adding Accountability and Educational Development

The prior analysis presumes that a country's income level can sufficiently characterize the set of institutional features that are complementary to local autonomy in schools – including, for example, experience with general economic structures, the importance of the rule of law as seen in economic operations, generally functioning governmental institutions, and the like. It has the potential disadvantages of ignoring specific educational institutions and the overall development of the educational sector. For these reasons, we present exploratory estimates of more education-specific features of a country that might provide a more refined look at autonomy.

As described in our conceptual principal-agent framework, the effect of autonomy may not only depend on the level of development, but also on the extent to which a school system directly monitors results through accountability systems. Existing cross-sectional research has found significant interactions of school-level autonomy with country-level existence of the accountability measure of central exit exams across countries (see Woessmann, 2005; Hanushek and Woessmann, 2011b). Thus, the first column of Table 2.9 adds an interaction term between autonomy and central exit exams to our basic model. There is a sizeable positive interaction between (time-variant) school autonomy and the (time-invariant) measure of central exit exams, statistically significant in the case of academic-content autonomy. The effect of introducing autonomy is more positive in countries that hold the system accountable by central exit exams. At the same time, our main effect of an interaction between autonomy and level of development is unaffected by including the autonomy-exam interaction. As is evident in column (2), there is no significant triple interaction between autonomy, exams, and initial GDP per capita, suggesting that the impact of the development level on the autonomy effect does not depend on whether there are central exams in the school system, and vice versa.

We have consistently measured the initial level of development by overall economic development (GDP per capita). An alternative way of measuring development is to look at the achievement level of the education system, which we measure by the initial average PISA score in 2000. As shown in Table 2.10, the effect of school autonomy indeed increases significantly with the initial achievement level. The negative autonomy effect in poorly performing systems is again larger for academic-content autonomy than for personnel autonomy. For a country at the relatively low initial achievement level of 400 PISA points, equivalent to one standard deviation below the OECD mean, going from no to full school autonomy reduces student achievement by 0.63 standard deviations in academic-content autonomy and by 0.33 standard deviations in personnel autonomy. The coefficient estimates imply that the autonomy effect turns from negative to positive at a performance level of 485 and 449 PISA points, respectively, for academic-content and for personnel autonomy. At the level of the highest-performing country (Hong Kong with a test score of 560.5), the positive effect of academic-content autonomy is as large as 0.56 standard deviations, and 0.72 standard deviations for personnel autonomy.

Column (2) of Table 2.10 jointly enters the interactions of autonomy with the initial PISA score and with initial GDP per capita. Both retain statistical significance for interactions with academic-content autonomy, while limited statistical power has the two interaction terms shy of statistical significance for personnel autonomy. Initial educational achievement and initial GDP per capita may thus capture two separable dimensions of the performance level of a country that have relevance for how school autonomy affects student outcomes.[33]

For robustness, the final two columns use alternative forms of measuring initial achievement. In column (3), qualitative results are similar when the initial achievement level is not measured linearly but as a dummy for countries scoring higher than 400 PISA points (one standard deviation below the OECD mean). Similarly, results hold when

[33] Results are robust to dropping the former Communist countries, which – as seen in Figure 2.3 – are noteworthy outliers in the plot of initial GDP per capita against initial achievement.

Table 2.9

Extended Model: Including Central Exit Exams

	(1)	(2)
Academic-content autonomy	-48.511**	-48.645**
	(19.363)	(19.921)
Academic-content autonomy x Central exit exams (CEE)	32.750**	32.931*
	(14.374)	(16.382)
Academic-content autonomy x Initial GDP p.c.	3.141***	3.168***
	(0.563)	(0.938)
Academic-content autonomy x CEE x Initial GDP p.c.		-0.042
		(1.161)
R^2	0.380	0.380
Personnel autonomy	-28.555*	-19.300
	(14.574)	(17.994)
Personnel autonomy x Central exit exams (CEE)	18.310	5.755
	(21.815)	(27.312)
Personnel autonomy x Initial GDP p.c.	3.446***	0.897
	(1.057)	(2.149)
Personnel autonomy x CEE x Initial GDP p.c.		3.493
		(2.545)
R^2	0.379	0.379

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,028,970 students, 41 countries, 152 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.10

Alternative Measure of Development Level: Initial Level of Student Achievement

Measure of initial achievement:	Average PISA score		Dummy for average PISA score above	
			400 points	500 points
	(1)	(2)	(3)	(4)
Academic-content autonomy	-63.257***	-60.480***	-85.567***	-32.530**
	(14.544)	(13.773)	(19.203)	(14.818)
Academic-content autonomy *x* Initial achievement	0.744***	0.601***	74.590***	73.258***
	(0.076)	(0.089)	(16.739)	(12.443)
Academic-content autonomy *x* Initial GDP p.c.		1.193**		
		(0.535)		
R^2	0.386	0.386	0.385	0.385
Personnel autonomy	-32.691*	-29.342	-51.538	-14.953
	(17.660)	(18.356)	(31.462)	(12.657)
Personnel autonomy *x* Initial achievement	0.654***	0.372	63.266*	82.534***
	(0.216)	(0.292)	(32.166)	(24.584)
Personnel autonomy *x* Initial GDP p.c.		1.991		
		(1.406)		
R^2	0.384	0.384	0.384	0.384

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the first two columns, the initial average PISA score is centered at 400 (one standard deviation below the OECD mean), so that the main effect shows the effect of autonomy on test scores in a country that in 2000 performed at a level one standard deviation below the OECD mean. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

measuring initial achievement by a dummy for countries scoring higher than the OECD mean of 500 PISA points (column 4). Results are also very similar for a dummy for countries above the sample median of 480 PISA scores (not shown).

We find both of these extensions – accountability and development of the educational system per se – to be highly suggestive of a more nuanced view of autonomy. At the same time, the limitations of our cross-country approach that come from relatively small effective samples of countries and from imperfect measurement of specific institutions lead us to be cautious in the interpretation. We think there are conceptual reasons that lend credence to these results, particularly about accountability, but there are many details about the form and consequences of accountability that are ignored.[34]

2.7 Conclusions

Decentralization of decision-making has been hotly debated in many countries of the world, and prior research has left considerable uncertainty about the expected impact of giving more autonomy to schools. In the face of this uncertainty, many countries have changed the locus of decision-making within their countries over the past decade – and interestingly some have decentralized while others have centralized. We exploit this cross-country variation to investigate the impact of local autonomy on student achievement. We identify the effect of school autonomy from within-country changes in the share of autonomous schools over time in a panel analysis with country (and time) fixed effects.

Our central findings are consistent with the interpretation that autonomy reforms improve student achievement in developed countries, but undermine it in developing countries. At low levels of economic development, increased autonomy actually appears to hurt student outcomes, in particular in decision-making areas related to academic content. By contrast, in high-income countries, increased autonomy over academic content, personnel, and budgets exerts positive effects on student achievement. In general, the autonomy effects are most pronounced in decision-making on academic content, with some additional relevance for personnel autonomy and, less so, for budgetary autonomy.

[34] To illustrate the details on accountability, see the alternative estimates of its impact on student achievement in the U.S. (Figlio and Loeb, 2011).

Empirically, the main result proves highly robust across a series of sensitivity and specification checks. Among others, the autonomy effects show up in various forms of measuring initial GDP per capita, alternative specifications of the control model, and different sub-samples in terms of included waves and countries. The basic finding of heterogeneity of the impact of autonomy by development level shows up in students performance in math, in reading, and in science. It is much more pronounced for full school-level autonomy than for joint authority between schools and external authorities.

In terms of the model specification, we confirm that policy decisions about the introduction of autonomy reforms are not related to previous levels of achievement and GDP per capita, corroborating the panel identification. In addition, there are no significant interactions of the development level with other education policy measures, suggesting that the specific institutional effect and its heterogeneity are particular to autonomy reforms. Also, the significant interaction of autonomy with the level of economic development prevails when interactions of autonomy with measures of democracy, governance effectiveness, cultural values, and effective school environments are additionally taken into account, and the latter interactions are not significantly related to student outcomes once the interaction with economic development is held constant. Finally, there is no indication that autonomy differentially affects students with well-off and disadvantaged backgrounds. This suggests that autonomy reforms do not affect inequality between students with different social backgrounds in either developed or developing countries.

There is an indication that local decision-making works better when there is also external accountability that limits any opportunistic behavior of schools. Further, having generally well-functioning schools, indicated by initial performance levels, appears complementary with autonomy. In contrast to the observed dimensions of general governance, cultures, and social backgrounds, levels of accountability and effectiveness of the education system may thus constitute relevant channels through which the level of economic development affects the effectiveness of autonomy policies. Nonetheless, these specific issues require further research and confirmation.

From an analytical perspective, the innovation in this work is the development of panel data that permit cross-country analysis. Within this framework, we can exploit the pattern of policy changes within countries to obtain cleaner estimates of the institutional differences.

Does school autonomy make sense everywhere? Our results suggest that the answer is a clear "no": The impact of school autonomy on student achievement is highly heterogeneous, varying by the level of development of a country. This overall result may have broader implications for the generalizability of findings across countries and education systems. It suggests that lessons from educational policies in developed countries may not translate directly into advice for developing countries, and vice versa.

At the same time, it is appropriate to close with a caution. Identifying causal impacts in cross-country analyses is inherently difficult (see Hanushek and Woessmann, 2011b). Obviously, in a variety of evaluations within countries, the identification of the key policy parameters is clearer. But, we view this as an important complement to rigorous within-country evaluations, because it is often very difficult to know how to generalize those results to different decentralization policies or to different countries. Indeed, it is also the case that some country policies cannot be readily evaluated within individual countries, for example, when the policies are applied simultaneously to all schools or when there are substantial general equilibrium effects. Yet, there is always a possibility that our estimates have been contaminated by other, correlated factors or policies. We have clearly eliminated some major factors – importantly, time-invariant cultural, institutional, and population differences. We have also provided a series of robustness and specification tests based on measured aspects of schools and countries. All consistently suggest a powerful and significant impact of autonomy but one that varies in efficacy across countries at different levels of development. While our precise estimates may be affected by further, unmeasured influences, we believe that the overall qualitative patterns are almost certainly real and ones that should enter into the policy discussions.

Table A2.1
Descriptive Statistics and Complete Model of Basic Specification

	Descriptive Statistics			Basic Model	
	Mean	SD	Share imp.	Coeff.	SE
Academic-content autonomy	0.663	(0.259)	-	-34.018***	(12.211)
Academic-content autonomy *x* Initial GDP p.c.	5.760	(8.512)	-	2.944***	(0.590)
Student and family characteristics					
Female	0.501		0.002	-13.028***	(0.917)
Age (years)	15.762	(0.300)	0.002	13.449***	(1.335)
Immigration background					
Native student	0.914		0.022		
First generation student	0.042		0.022	-20.976***	(4.690)
Non-native student	0.043		0.022	-12.607**	(5.124)
Other language/ dialect than test language	0.094		0.043	-9.181**	(3.692)
Parents' education					
None	0.022		0.036		
Primary	0.077		0.036	10.697***	(2.115)
Lower secondary	0.104		0.036	11.724***	(2.610)
Upper secondary I	0.090		0.036	20.863***	(3.381)
Upper secondary II	0.279		0.036	25.784***	(2.866)
University	0.429		0.036	32.766***	(3.019)
Parents' occupation					
Blue collar low skilled	0.116		0.043		
Blue collar high skilled	0.153		0.043	6.013***	(1.184)
White collar low skilled	0.230		0.043	14.502***	(1.155)
White collar high skilled	0.502		0.043	35.714***	(1.953)
Books at home					
0-10 books	0.140		0.026		
11-100 books	0.471		0.026	29.430***	(2.339)
101-500 books	0.307		0.026	63.003***	(2.650)
More than 500 books	0.082		0.026	74.589***	(3.329)
School characteristics					
Number of students	784	(596)	0.062	0.016***	(0.003)
Privately operated	0.192		0.070	6.438	(4.481)
Share of government funding	0.841	(0.521)	0.079	-18.628***	(5.153)
Share of fully certified teachers at school	0.777	(0.330)	0.232	15.669***	(3.786)
Shortage of math teachers	0.183		0.027	6.984***	(1.449)
School's community location					
Village or rural area (<3,000)	0.110		0.046		
Town (3,000-15,000)	0.210		0.046	4.816**	(2.220)
Large town (15,000-100,000)	0.322		0.046	8.097***	(2.563)
City (100,000-1,000,000)	0.220		0.046	11.182***	(3.016)
Large city (>1,000,000)	0.138		0.046	12.191***	(3.633)
GDP per capita ($1,000)	24.973	(19.311)	-	0.416*	(0.245)
Country fixed effects				Yes	
Year fixed effects				Yes	
Student observations	1,042,995			1,042,995	
Country observations	42			42	
Country-by-wave observations	155			155	
R^2				0.385	

Notes: Descriptive statistics: Mean: international mean (weighted by sampling probabilities). SD: international standard deviation (only for continuous variables). Share imputed: share of missing values in the original data, imputed in the analysis. SE: robust standard errors. Basic model: Full results of the specification reported in the top panel of Table 4. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability. Regression includes imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.2
Questionnaire Item on Autonomy Across PISA Waves

Wave	Question	Answer options
2000	In your school, who has the main responsibility for . . . (Please tick as many boxes as appropriate in each row)	Not a school responsibility Appointed or elected board Principal Department head Teachers
2003	In your school, who has the main responsibility for . . . (Please tick as many boxes as appropriate in each row)	Not a main responsibility of the school School's governing board Principal Department head Teacher(s)
2006	Regarding your school, who has a considerable responsibility for the following tasks? (Please tick as many boxes as appropriate in each row)	Principals or teachers School governing board Regional or local education authority National education authority
2009	Regarding your school, who has a considerable responsibility for the following tasks? (Please tick as many boxes as appropriate in each row)	Principals Teachers School's governing board Regional or local education authority National education authority

Notes: For each decision-making task, we constructed a variable indicating full autonomy at the school level if the principal, the school's board, department heads, or teachers carry sole responsibility. Consequently, if the task is not a school responsibility (2000 and 2003 data) or the responsibility is also carried at regional/local or national education authorities (2006 and 2009 data), we do not classify a school as autonomous. Figure 2 does not indicate consistent changes across waves in the measure of autonomy over countries or tasks, indicating that changes in response options are unlikely to substantially affect our estimates. Furthermore, in our models, time fixed effects capture consistent changes between waves.

Table A2.3
Robustness: Correlation Between EAG and PISA Autonomy Measures

	EAG autonomy measures in 1998				
	Combined	Instruction	Personnel	Structure	Resources
PISA measures in 2000					
Academic-content	0.5	0.52	0.35	0.12	0.27
	(0.02)	(0.02)	(0.13)	(0.60)	(0.23)
Personnel	0.45	0.27	0.35	0.14	0.46
	(0.04)	(0.23)	(0.12)	(0.56)	(0.04)
Budget	0.26	0.29	0.36	-0.06	0.15
	(0.25)	(0.2)	(0.11)	(0.79)	(0.52)
Countries	21	21	21	21	21

	EAG autonomy measures in 2003				
	Combined	Instruction	Personnel	Structure	Resources
PISA measures in 2003					
Academic-content	0.53	0.65	0.43	0.18	0.26
	(0.01)	(0.001)	(0.04)	(0.40)	(0.23)
Personnel	0.5	0.41	0.48	0.13	0.41
	(0.02)	(0.05)	(0.02)	(0.55)	(0.05)
Budget	0.13	-0.03	0.25	0.09	0.17
	(0.54)	(0.88)	(0.25)	(0.6ť8)	(0.43)
Countries	23	23	23	23	23

	EAG autonomy measures in 2007				
	Combined	Instruction	Personnel	Structure	Resources
PISA measures in 2006					
Academic-content	0.48	0.56	0.4	0.19	0.32
	(0.03)	(0.01)	(0.07)	(0.41)	(0.15)
Personnel	0.57	0.46	0.64	0.25	0.56
	(0.01)	(0.04)	(0.00)	(0.28)	(0.01)
Budget	0.26	0.29	0.36	-0.06	0.15
	(0.25)	(0.2)	(0.11)	(0.79)	(0.52)
Countries	21	21	21	21	21

Notes: Each cell presents the correlation between one of the PISA autonomy measures and one of the Education at a Glance (EAG) autonomy measures. Standard errors in parentheses.

Table A2.4
Disaggregation of Basic Model: Results for Separate Autonomy Categories

	Estimation result		Details on autonomy effect at different levels of GDP per capita		
	Main effect (at initial GDP p.c. of $8,000)	Interaction with initial GDP p.c.	GDP p.c. at which autonomy effect switches sign	Effect in country with minimum GDP p.c.	Effect in country with maximum GDP p.c
	(1)	(2)	(3)	(4)	(5)
School autonomy over courses	-21.753***	2.338***	17,304	-38.578***	47.153***
	(8.050)	(0.468)		(10.293)	(11.524)
R^2	0.385			0.385	0.385
School autonomy over content	-28.142***	2.278***	20,354	-44.538***	39.009***
	(10.257)	(0.509)		(12.880)	(11.680)
R^2	0.385			0.385	0.385
School autonomy over textbooks	-24.233***	2.675***	17,059	-43.483***	54.605**
	(8.124)	(0.772)		(11.511)	(20.937)
R^2	0.385			0.385	0.385
School autonomy over hiring	-32.973*	3.314***	17,950	-56.821**	64.698***
	(16.933)	(1.117)		(23.602)	(22.979)
R^2	0.384			0.384	0.384
School autonomy over salaries	-3.135	2.389**	9,312	-20.324	67.263**
	(12.305)	(1.110)		(16.138)	(32.188)
R^2	0.384			0.384	0.384
School autonomy over budget allocations	-6.347	1.838**	11,453	-19.576	47.833**
	(9.363)	(0.796)		(13.282)	(20.221)
R^2	0.384			0.384	0.384

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8000. "Maximum GDP p.c." refers to Norway. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.5

Robustness: Main Specification Without Controlling for GDP p.c., with GDP p.c. and its Changes and its Growth Rate

	Academic-content autonomy			Personnel autonomy			Budget autonomy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Autonomy	-28.418**	-31.634**	-31.830**	-18.107	-17.950	-14.924	-6.438	-7.544	-4.432
	(10.562)	(12.333)	(11.952)	(14.011)	(13.482)	(13.970)	(9.235)	(8.919)	(8.434)
Autonomy x Initial GDP p.c.	2.693***	2.875***	2.816***	3.269***	3.174***	2.947**	1.093	1.317	1.352*
	(0.539)	(0.589)	(0.568)	(0.932)	(1.078)	(1.103)	(0.984)	(0.869)	(0.759)
GDP p.c.		0.248	0.398		-0.213	-0.029		0.121	0.288
		(0.275)	(0.244)		(0.321)	(0.298)		(0.378)	(0.340)
Δ GDP p.c. (in $1,000)		0.457			0.550**			0.505*	
		(0.285)			(0.239)			(0.289)	
GDP growth rate			3.569			7.764*			7.899**
			(3.825)			(4.156)			(3.650)
R^2	0.385	0.385	0.385	0.384	0.385	0.384	0.384	0.384	0.384

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.6
Robustness: Main Results Excluding Private Schools

	Main effect (at initial GDP p.c. of $8,000)	Effect in country with minimum GDP p.c.	Effect in country with maximum GDP p.c
	(1)	(2)	(3)
Academic-content autonomy	-32.373**	-52.309***	49.276***
	(12.419)	(14.691)	(17.575)
Academic-content autonomy x Initial GDP p.c.	2.770***	2.770***	2.770***
	(0.614)	(0.614)	(0.614)
R^2	0.375	0.375	0.375
Personnel autonomy	-22.015*	-46.316***	77.513**
	(11.840)	(17.074)	(31.233)
Personnel autonomy x Initial GDP p.c.	3.377***	3.377***	3.377***
	(1.160)	(1.160)	(1.160)
R^2	0.374	0.374	0.374
Budget autonomy	-4.098	-16.959	48.578**
	(9.174)	(13.160)	(23.381)
Budget autonomy x Initial GDP p.c.	1.787**	1.787**	1.787**
	(0.875)	(0.875)	(0.875)
R^2	0.374	0.374	0.374

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. "Maximum GDP p.c." refers to Norway. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009 at public schools. Sample size in each specification: 795,260 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.7
Alternative Estimation of the Impact of Autonomy: Country-level Estimation of Two-step Model

	First step does not include country fixed effects	First step includes country fixed effects
	(1)	(2)
Academic-content autonomy	-30.247**	-26.378**
	(12.757)	(10.691)
Academic-content autonomy *x* Initial GDP p.c.	3.025***	2.892***
	(0.817)	(0.701)
R^2	0.869	0.186
Personnel autonomy	-8.219	-14.322
	(1.494)	(15.116)
Personnel autonomy *x* Initial GDP p.c.	3.172**	3.348***
	(17.284)	(1.257)
R^2	0.856	0.099
Budget autonomy	-8.700	-7.480
	(12.141)	(10.773)
Budget autonomy *x* Initial GDP p.c.	1.679	0.945
	(1.319)	(1.010)
R^2	0.853	0.051

Notes: Each panel-by-column presents results of a separate regression. Reported coefficients stem from a country-level least squares regression with country and year fixed effects, controlling for GDP per capita. Sample: country-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 155 country-by-wave observations covering a total of 42 countries. Dependent variable: Country-level aggregation of the residuals of a first-step estimation at the student level that regresses the PISA math test score on student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home, school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school, and imputation dummies (and, in column 2, country's GDP per capita, country fixed effects, and year fixed effects). Initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Robust standard errors in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.8
Robustness: Controlling for Enrollment Rates

	Estimation result	Details on autonomy effect at different levels of GDP per capita	
	Main effect (at initial GDP p.c. of $8,000)	Effect in country with minimum GDP p.c.	Effect in country with maximum GDP p.c
	(1)	(2)	(3)
Academic-content autonomy	-31.253***	-52.415***	55.420***
	(11.059)	(13.078)	(16.016)
Academic-content autonomy x Initial GDP p.c.	2.941***		
	(0.556)		
Enrollment Rate	50.001		
	(32.990)		
R^2	0.386		
Personnel autonomy	-15.000	-38.016**	79.261***
	(13.330)	(18.164)	(27.519)
Personnel autonomy x Initial GDP p. c.	3.198***		
	(1.055)		
Enrollment Rate	50.819		
	(36.408)		
R^2	0.385		
Budget autonomy	-12.501*	-28.859***	54.496***
	(7.378)	(10.102)	(17.740)
Budget autonomy x Initial GDP p. c.	2.273***		
	(0.649)		
Enrollment Rate	59.980		
	(36.218)		
R^2	0.385		

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. "Maximum GDP p.c." refers to Norway. Data on enrollment rates is taken from the respective PISA publications, missing for Romania in 2000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,040,313 students, 42 countries, 154 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A2.9

Robustness: Regressing Changes in Autonomy on Initial GDP per Capita

	Academic-content autonomy		Personnel autonomy		Budget autonomy	
Change between	2003-2000	2009-2000	2003-2000	2009-2000	2003-2000	2009-2000
	(1)	(2)	(3)	(4)	(5)	(6)
GDP pc in 2000	0.003	0.000	0.002	-0.001	-0.000	-0.004**
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Constant	-0.074	0.006	-0.040	0.032	0.031	0.109***
	(0.062)	(0.021)	(0.050)	(0.019)	(0.030)	(0.038)
Observations	29	38	29	38	29	38
R^2	0.041	0.001	0.017	0.047	0.002	0.133

Notes: Each panel presents results of a separate regression. Dependent variable: Change in autonomy variable between 2003 and 2000 and between 2009 and 2000. Sample: country-level observations in PISA waves 2000, 2003, 2006, and 2009. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A2.10

Robustness: Use Alternative EAG Measure in Main Specification

	A) EAG data: Share of Decisions taken at School with Full Autonomy				
	Combined	Instruction	Structure	Personnel	Resource
	(1)	(2)	(3)	(4)	(5)
Autonomy	-83.409**	-60.707**	-46.774**	-22.908	83.300
	(38.431)	(24.836)	(21.369)	(17.946)	(52.223)
Autonomy x Initial GDP p.c.	5.398*	3.837*	2.426	1.249	-4.985
	(2.702)	(1.971)	(1.876)	(0.726)	(3.228)
R^2	0.302	0.302	0.302	0.302	0.302

	B) As Comparison: PISA Measure Using Same Sample		
	Academic-content	Personnel	Budget
	(1)	(2)	(3)
Autonomy	-10.544	-16.063	-48.552***
	(9.391)	(14.666)	(16.791)
Autonomy x Initial GDP p.c.	2.011**	1.949	3.244**
	(0.766)	(1.427)	(1.293)
R^2	0.302	0.302	0.302

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. In the upper part of this table, school autonomy is measured as share of decisions at school with full autonomy reported by the Education at a Glance (EAG) volumes. In the lower part of this table, school autonomy is measured as country average. Initial GDP per capita is centered at $8,000 (measured in $1,000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of $8,000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 373,424 students, 22 countries, 57 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Chapter 3

The Effect of Single-Sex Schooling on Student Performance: Quasi-Experimental Evidence from South Korea

3.1 Introduction

Underrepresentation of women in both high-paying and STEM-related occupations is well documented (Organisation for Economic Co-operation and Development, 2008b).[1] Explanations include gender-stereotype sorting, differences in individual preferences, in non-cognitive behavior or in cognitive skills between men and women. Differences in the distribution of quantitative skills between boys and girls partly explain the sorting of men and women into high-paying and low-paying fields (Paglin and Rufolo, 1990).[2] However, variation across cultures suggests that this gap is due to the social environment rather than inherent gender traits (Guiso et al., 2008; Fryer and Levitt, 2010).[3] Thus, raising girls' interest and achievement in math and sciences is a goal of policy aimed

[1] STEM is an acronym for the fields of study of science, technology, engineering, and mathematics.

[2] Paglin and Rufolo (1990) show that there is a much higher proportion of men than women in the top intervals of mathematical reasoning ability, which is often a qualification in high-paying fields. Interestingly, women with high mathematical reasoning abilities also show high participation rates in STEM-related occupations.

[3] Booth et al. (2011) emphasize the importance of social learning rather than inherent gender traits for observed gender differences in risk behavior.

at reducing gender-based disparities. In this context, single-sex schooling has gained particular attention. For example, in the United States, single-sex classrooms are a growing phenomenon and amendments to Title IX that explicitly allow single-sex public schools and classes have set off a pedagogical dispute over whether sex-segregation improves educational achievement (Cohen, 2012; Whitmore, 2005).[4]

Although arguments for and against single-sex education are well-developed, the underlying mechanisms of the effect of single-sex schooling are not well understood. Most pronounced in the public debate is the argument that the presence of the opposite sex in the classroom is distracting and leads to lower educational achievement for both boys and girls (Coleman, 1961). In line with this, single-sex schools are claimed to have more serious and studious classroom climates (Lee and Bryk, 1986). This might be especially beneficial for girls given that boys are more disruptive, restless, and dominant in class. In fact, larger shares of girls in class are found to be associated with higher academic achievement which can partly be explained by a lower level of classroom disruption and violence (Hoxby, 2000; Whitmore, 2005; Lavy and Schlosser, 2011).[5] Other explanations for positive effects of single-sex schooling include gender-tailored teaching styles and more positive attitudes towards stereotypically male subjects for girls at single-sex schools.

This paper investigates the effect of single-sex schooling on academic achievement in two stereotypically male subjects, namely, math and science, at Korean middle schools. As a result of unequal education opportunities and the clustering of students with high socio-economic background at elite schools, the Korean government introduced a lottery system to allocate students to schools, regardless whether they are coeducational or single-sex organized. This ensures that attendance at single-sex schools is orthogonal to student characteristics such as socio-economic background and ability, such that the comparison between girls (boys) at coeducational schools and girls (boys) at single-sex schools should identify a reliable effect of single-sex schooling on student achievement. Moreover, the rich dataset used in this study allows us to observe a large number of qualitative indicators, such as student attitude and teaching style, to investigate a broad set of potential channels.

[4] For example, Billger (2009) studies the effects of single-sex schooling in the context of the increase in single-sex classes and schooling in the United States as a response to amendments to Title IX.

[5] However, Whitmore (2005) reports positive effects for boys only until second grade. In third grade, boys do actually worse if they are in a class with a high fraction of girls.

We find positive effects of single-sex schooling for girls in math which are highly statistically significant and non-negligible in their magnitude. This finding is particular relevant since math performance is consistently linked to future earnings (Paglin and Rufolo, 1990). For science, the coefficients show a similar pattern, but are generally smaller and not significant. In contrast, we do not find beneficial or adverse effects for boys.[6] Subgroup analyses show that especially girls with non-supporting parental background benefit from single-sex schooling, with the effects being largest among the group of low-performing students. Moreover, the comparison across gender reveals an interesting pattern regarding the well-known test score gender gap in math. While there is no test score gap in math between girls and boys with high socio-economic backgrounds, regardless whether they attend coeducational or single-sex schools, girls from low socio-economic backgrounds at coeducational schools fall behind their male classmates and their female peers at single-sex schools. The exploration of potential channels shows that these effects can neither be explained by differences in school and teacher characteristics at coeducational and single-sex schools nor by gender-tailored teaching practices or more positive attitudes toward math at single-sex schools. However, some of the effect can be attributed to rougher classroom atmospheres at mixed schools. Several robustness tests suggest that the results are not driven by differences in the types of students that attend single-sex and coeducational schools.

Despite a great deal of work on the subject, empirical evidence regarding the effects of single-sex schooling on student outcomes is inconclusive (Bigler and Signorella, 2011). Several studies report positive effects, especially for girls, on academic achievement, self-esteem, and other non-cognitive outcomes (e.g., Lee and Bryk, 1986; Riordan, 1990; Jackson, 2002). However, other studies find no significant differences between students at coeducational and single-sex schools (Marsh, 1989). Moreover, most of the literature is based on the comparison of student outcomes at coeducational schools and single-sex schools. These results are likely to be biased by self-selection of students into single-sex schools, since attendance at single-sex-schools is usually correlated with unobservable, individual characteristics that also determine student achievement.[7]

[6] This finding is in line with earlier research on the effects of single-sex education. For example, Jackson (2002) finds positive effects of all-girl classes, but no effects for all-boy classes.

[7] See, e.g., Lee and Bryk (1986) for an overview of reasons to choose single-sex schools over coeducational schools.

In recent years, there has been a growing literature that addresses these selection issues as to isolate the causal effect of single-sex schooling on student outcomes. Jackson (2012) exploits the fact that assignment rules in Trinidad and Tobago create exogenous variation to remove selection bias. He shows that only girls with stated preferences for single-sex schooling actually perform better. However, for most students he finds no significant effects. Eisenkopf et al. (2011) report positive effects of single-sex education for girls at a Swiss high school where girls are randomly assigned to single-sex and coeducational classrooms. In a similar manner, Behrman et al. (2012) make use of a unique feature in the Korean education system, namely, the random allocation of students to high schools in Seoul. They show that attending a single-sex school is associated with higher test scores in Korean and English and a higher probability of attending a four-year college for both girls and boys.

This chapter contributes to the growing quasi-experimental literature on the effect of single-sex schooling by following Behrman et al. (2012). We focus on Korean middle schools which are compulsory and therefore represent the full population of students and investigate the effect of single-sex schooling on math and science which is especially interesting given the discussion about the influence of gender stereotypes on student achievement and choice (Thompson, 2003; Joshi et al., 2010; Favara, 2011). Moreover, given that the underlying mechanisms of single-sex education are not well understood, the exploration of a broad set of potential channels is an important contribution to the literature. The results suggest that girls with non-supporting family backgrounds are harmed by a rougher atmosphere at coeducational schools. The more general implication may be that in any school system, girls with a non-supporting background may be particularly influenced by less favorable peer characteristics.[8]

The remainder of this chapter is organized as follows: Section 3.2 describes the random assignment process to Korean middle schools. Section 3.3 presents the data. Section 3.4 explains the empirical approach. Section 3.5 provides the main results. Section 3.6 explores potential channels and mechanisms. Section 3.7 investigates hheterogeneous effects of single-sex schooling. Section 3.8 reports the results of several robustness tests. Section 3.9 summarizes and concludes.

---

[8] The heterogeneity of peer effects across gender is also documented by Lavy et al. (2012), who show that only girls significantly benefit from the presence of academically strong peers.

3.2 The Random Assignment Process

As a response to very low enrollment rates after the Japanese liberalization, elementary schooling in Korea became universal and compulsory in 1951. Despite limited school facilities and resources after the 3-year Korean War, enrollment rates for elementary schooling increased remarkably and rose steadily (Kim and Lee, 2003). Since most resources were invested in the primary education sector, the capacity of public secondary schools was not much increased. As a result, the provision of secondary school facilities lagged behind the rapid growth of the student population and resulted in a fierce competition among students in the admission process to middle and high schools. Consequently, all middle and high schools selected their students through competitive entrance examinations. However, the selection of students based on entrance examinations resulted in an advantage for wealthy families that were able to better support their children, particularly by paying for private tutoring. At that time, the Korean education system was characterized by an excess demand for secondary schools, substantial quality differences across schools, and overall unequal education opportunities.

As a response, an "Equalization Policy" (EP) was introduced for middle schools in 1969 with the aim of creating equal education opportunities and reducing the influence of social background on student educational achievement.[9] Under this policy, the competitive entrance examinations were replaced by a random allocation, via a lottery system, of middle school students within each school district. In other words, all middle schools, regardless of whether they were public or private, could no longer select students themselves but instead were required to take all students assigned to them by the Ministry of Education via a district-wide lottery. Moreover, the policy required equalization of school resources and teachers in an effort to ensure that there were no differences in resources and instruction quality across schools (Kim and Lee, 2003). Curriculum and teacher qualifications became uniform and centrally regulated. The government even provided subsidies to financially weak private schools so that their teacher salaries would equal those of public schools.[10]

---

[9] See, e.g., Organisation for Economic Co-operation and Development (1998); Kim and Lee (2003); Ju-Ho (2004); Behrman et al. (2012) for an overview of the Korean education system in general and the education reforms in particular.

[10] Since public provision of secondary schools did not meet the demand, the gap was filled up by an increasing number of private schools. However, for-profit secondary schools were permitted and tuition fees were under government regulation. Since private schools received no additional government funding,

Furthermore, all private schools were required to charge the same tuition and teach the central curriculum.[11]

The policy was first implemented to middle schools in Seoul in 1969 and expanded to major cities and then throughout the entire country within the next two years. Differences in teacher quality and school resources between schools were quickly reduced and improvements in the physical and psychological development of children reported. However, now that the problem was solved at the middle school level, an even fiercer competition for prestigious high schools began. As a response, the government introduced the “High School Equalization Policy” in 1974 for general high schools. Under this policy, entrance examinations for general high schools were abolished. After passing a screening process, applicants for general high schools were assigned by lottery to a school within their residential district. Again, the policy was first adopted in Seoul and Pusan, the two largest Korean cities. By 1980, the Equalization Policy had been expanded to cover most major Korean cities.

The original structure of the Equalization Policy has been maintained for the past 30 years, leaving its main guidelines unchanged.[12] Even today, all middle school students are assigned by lottery to a school within their residential district (Kim and Lee, 2003; Ju-Ho, 2004). However, the High School Equalization Policy became the subject of discussion and critique during the 1990s. As a result, its implementation was slowed. Metropolitan cities continued to be required to follow the policy and assign their students to general high schools, but it was optional for smaller cities and rural school districts. More recently, some school districts modified the High School Equalization Policy such that students are allowed to state their preferences and high schools may choose a fraction of their students.

3.3 Data

Secondary schooling in Korea is organized into lower and upper secondary education. After graduating from middle school, which ranges from Grade 7 to 9, students usually proceed

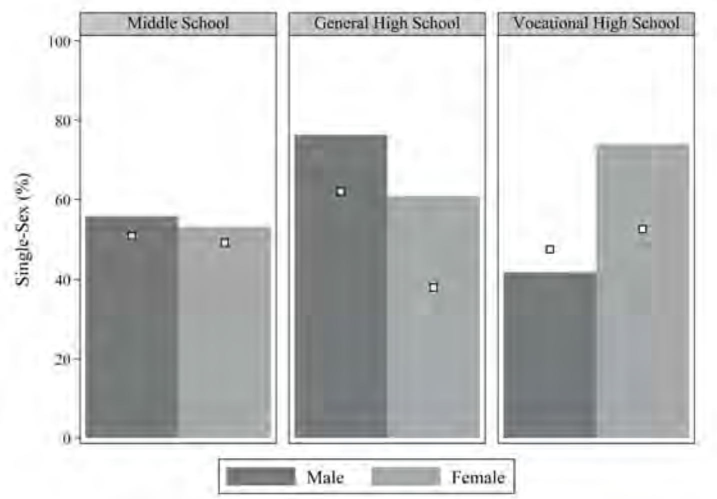
as did public schools, private schools were generally less well-funded and had inferior facilities compared to public schools.

[11] Except for certain rights over personnel decisions and school facilities, there are almost no differences between private and public schools in Korea. Even essential features of private schools, such as selection of students, tuition fees, teacher salaries, and curriculum, are regulated.

[12] See Ju-Ho (2004) for a debate on the High School Equalization Policy.

to upper secondary education and attend either general or vocational high schools.[13] Figure 3.1 presents the Korean student population by school type and gender. The proportions of girls and boys at middle schools are quite equal – presumably because middle school is compulsory – whereas the share of boys exceeds the share of girls at general high schools which are more academically oriented (denoted by white squares). More importantly, this figure reveals Korea's long-standing tradition of single-sex schooling and shows that students are nearly equally divided between single-sex and coeducational schools (denoted by bars).

Figure 3.1. Student Population by School Type and Gender.



Notes: The white squares denote the overall share of girls and boys by type of school. The bars denote the shares of girls and boys that attend single-sex schools. Data source: TIMSS 1999 for middle school students, PISA 2000 for high school students.

Given that we are interested in the effect of single-sex schooling on student achievement, we focus on middle schools only because these are covered by the Equalization Policy without any exceptions. In contrast, vocational high schools have always been excluded from the Equalization Policy and, as outlined above, general high schools are subject to a number of exceptions.[14] Since the Equalization Policy was not modified subsequent to

[13] Graduates of vocational high schools are qualified for direct entry into the labor market. In contrast, general high schools are more academically oriented and qualify their graduates for tertiary education. In addition, there are a few specialized schools.

[14] The results for high school students in areas that are most likely be covered by the Equalization Policy are provided in the Appendix.

the emergence of coeducational schools, middle school students are randomly assigned to schools, regardless whether they are coeducational or single-sex organized.

We use the Third International Mathematics and Science Study (TIMSS) that provides educational achievement indicators and extensive background information for eighth grade students as well as information on school and teacher characteristics. To indicate the single-sex status of a middle school, we rely on information as to the number of girls and boys enrolled at a school which is most recently collected in TIMSS 1999.[15] We drop observations from villages or rural areas because those areas are likely to have only a limited number of schools to which students could be assigned. In other words, by restricting the sample to large towns and cities, we focus on areas where the average school district has several coeducational as well as several all-girl and all-boy schools. For instance, in a typical school district (Kangnam) within the capital of Seoul, there are ten coeducational schools, seven all-boy schools and seven all-girl schools to which students can be assigned (Behrman et al., 2012). The resulting dataset totals 4,775 individual observations at 116 middle schools.

Tables 3.1 and 3.2 report student characteristics separately by gender for students at single-sex and coeducational middle schools. Since students are randomly assigned to middle schools, student characteristics should not differ across single-sex and coeducational schools. Table 3.1 reports conventional background characteristics such as students' age, parent's education and the number of books at home which is a proxy for socio-economic background. Strikingly, the share of female students with parents with secondary education is about eight percent larger at single-sex schools. However, this is actually compensated by a seven percent larger share of students at coeducational schools with parents holding a university degree. A similar pattern can also be reported among boys. In contrast, there are almost no differences for the number of books at home. In addition, Table 3.2 reports a large set of variables on home resources and on indicators of how students spent their time at home. Boys at single-sex schools are less likely to have a computer and internet at home and there are some differences in the frequency of watching news for girls. Overall, however, the random assignment of students to schools is reflected in very few significant differences between students at coeducational and single-sex schools.

[15] Since there are also single-sex classes at coeducational schools, we do not infer single-sex school status by the share of girls in a class in more recent waves of TIMSS. Single-sex classes at coeducational schools are further investigated in Section 3.5.2.

Table 3.1
Descriptive Statistics on Student Characteristics I

	Female						Male					
	Coed		Single-Sex		Diff.		Coed		Single-Sex		Diff.	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	14.45	(0.32)	14.42	(0.33)	-0.03**	(0.01)	14.42	(0.34)	14.43	(0.35)	0.01	(0.01)
Parents' education												
None	0.04	(0.19)	0.04	(0.20)	0.00	(0.01)	0.03	(0.17)	0.05	(0.21)	0.02**	(0.01)
Primary	0.13	(0.34)	0.13	(0.34)	0.00	(0.01)	0.12	(0.32)	0.13	(0.34)	0.02	(0.01)
Secondary	0.47	(0.50)	0.55	(0.50)	0.08***	(0.02)	0.42	(0.49)	0.47	(0.50)	0.05**	(0.02)
University	0.29	(0.45)	0.22	(0.41)	-0.07***	(0.02)	0.35	(0.48)	0.25	(0.43)	-0.10***	(0.02)
Books at home												
0-10 Books	0.08	(0.27)	0.08	(0.27)	-0.00	(0.01)	0.08	(0.28)	0.11	(0.32)	0.03***	(0.01)
11-25 Books	0.11	(0.31)	0.09	(0.29)	-0.02	(0.01)	0.09	(0.28)	0.11	(0.31)	0.02	(0.01)
26-100 Books	0.35	(0.48)	0.37	(0.48)	0.02	(0.02)	0.35	(0.48)	0.35	(0.48)	-0.00	(0.02)
101-200 Books	0.25	(0.43)	0.24	(0.43)	-0.00	(0.02)	0.24	(0.43)	0.22	(0.42)	-0.01	(0.02)
> 200 Books	0.21	(0.41)	0.21	(0.41)	0.01	(0.02)	0.24	(0.42)	0.21	(0.41)	-0.03*	(0.02)
Live with Parents	0.90	(0.30)	0.90	(0.30)	-0.00	(0.01)	0.90	(0.30)	0.90	(0.30)	-0.01	(0.01)
Observations	1122		1226		2348		1092		1335		2427	

Notes: Data source: TIMSS 1999. Individual observations are weighted by sampling probabilities. Standard deviations in parentheses. The category "do not know" is omitted for parent's education. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.2
Descriptive Statistics on Student Characteristics II

	Female						Male					
	Coed		Single-Sex		Diff.		Coed		Single-Sex		Diff.	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Home resources												
Computer at home	0.63	(0.48)	0.63	(0.48)	0.00	(0.02)	0.78	(0.42)	0.68	(0.47)	-0.09***	(0.02)
Internet at home	0.20	(0.40)	0.22	(0.41)	0.01	(0.02)	0.32	(0.47)	0.23	(0.42)	-0.10***	(0.02)
Calculator at home	0.95	(0.22)	0.96	(0.19)	0.02*	(0.01)	0.97	(0.18)	0.96	(0.18)	-0.00	(0.01)
Read a book												
about every day	0.22	(0.41)	0.23	(0.42)	0.00	(0.02)	0.28	(0.45)	0.24	(0.43)	-0.03	(0.02)
about once a week	0.44	(0.50)	0.43	(0.49)	-0.01	(0.02)	0.38	(0.49)	0.38	(0.49)	0.00	(0.02)
rarely/once a month	0.34	(0.47)	0.35	(0.48)	0.01	(0.02)	0.34	(0.48)	0.37	(0.48)	0.03	(0.02)
Watch news or documentaries												
about every day	0.27	(0.44)	0.31	(0.46)	0.04**	(0.02)	0.30	(0.46)	0.30	(0.46)	0.00	(0.02)
about once a week	0.33	(0.47)	0.35	(0.48)	0.03	(0.02)	0.34	(0.47)	0.36	(0.48)	0.02	(0.02)
rarely/once a month	0.41	(0.49)	0.34	(0.47)	-0.06***	(0.02)	0.36	(0.48)	0.34	(0.47)	-0.03	(0.02)
Go to the movies												
about every day/ once a week	0.03	(0.18)	0.03	(0.18)	0.00	(0.01)	0.04	(0.20)	0.06	(0.23)	0.02*	(0.01)
about once a month	0.46	(0.50)	0.46	(0.50)	0.00	(0.02)	0.45	(0.50)	0.44	(0.50)	-0.01	(0.02)
rarely	0.50	(0.50)	0.51	(0.50)	0.01	(0.02)	0.51	(0.50)	0.50	(0.50)	-0.01	(0.02)
Watch opera, ballet, classic music												
about every day/ once a week	0.06	(0.24)	0.06	(0.24)	-0.00	(0.01)	0.05	(0.22)	0.05	(0.23)	-0.01	(0.00)
about once a month	0.18	(0.38)	0.17	(0.38)	-0.00	(0.02)	0.12	(0.32)	0.14	(0.35)	0.03**	(0.01)
rarely	0.76	(0.43)	0.76	(0.43)	0.00	(0.02)	0.83	(0.37)	0.80	(0.40)	-0.03*	(0.02)
Watch comedies												
about every day	0.53	(0.50)	0.52	(0.50)	-0.01	(0.02)	0.54	(0.50)	0.54	(0.50)	-0.01	(0.02)
about once a week	0.34	(0.47)	0.35	(0.48)	0.01	(0.02)	0.34	(0.47)	0.34	(0.47)	0.00	(0.02)
rarely/about once a month	0.13	(0.34)	0.13	(0.33)	0.00	(0.01)	0.12	(0.32)	0.12	(0.33)	0.00	(0.01)
Observations	1112		1121		2233		1067		1318		2385	

Notes: Data source: TIMSS 1999. Individual observations are weighted by sampling probabilities. Standard deviations in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

3.4 Identification Strategy and Empirical Model

In the literature, the effect of attending a single-sex school is often derived by comparing students at coeducational and single-sex schools while controlling for a rich set of background variables. However, these estimates are unbiased only if the variable of interest, attendance at a single-sex school, is not correlated with unobserved characteristics captured by the error term. To satisfy this assumption, recently a number of studies make use of quasi-experimental settings (Eisenkopf et al., 2011; Behrman et al., 2012; Jackson, 2012).

To obtain the effect of single-sex schooling on student performance, we estimate the following model:

$$T_{ic} = \alpha + \beta SS_c + \gamma' X_{ic} + \varepsilon_i + \eta_c. \tag{3.1}$$

T_{ic} is student i's test score at school c in either math or science, while SS_c indicates if student i is attending a single-sex school (1, if single-sex). The dependent variable is normalized with a mean of 0 and a standard deviation of 1. X_{ic} denotes a large set of control variables at the individual, school, and teacher level, ε_i represents an idiosyncratic error term, and η_c the error component that varies at the school level. In all regressions, we cluster standard errors at the school level to account for the fact that students at the same school share similar background and identical school and teacher characteristics.[16]

As mentioned above, the causal interpretation of β relies on the underlying assumption that attendance at single-sex schools is orthogonal to unobserved individual characteristics. Given that students are randomly assigned to Korean middle schools, regardless whether they are coeducational or single-sex organized, this assumption no longer seems particularly strong. Tables 3.1 and 3.2 confirm, except for a few variables, that students at single-sex and coeducational schools have very similar observable characteristics.

Unfortunately, we cannot observe school districts. Thus, the few reported differences in student characteristics reported in Tables 3.1 and 3.2 could reflect differences in residential populations across school locations. However, this might only bias the estimates if the distribution of single-sex and coeducational schools is systematically related to

[16] Throughout the chapter, we report estimation results using the first plausible value reported in the data. However, the results are robust to using other plausible values.

the characteristics of school districts. In other words, if all-girl schools are located in better-off areas and coeducational schools in disadvantaged areas, any effects we find could be attributed to differences in the student composition at single-sex and coeducational schools. To account for this, we control for peer quality at the class level as a proxy for neighborhood characteristics.[17]

We run all regressions separately for girls and boys, thus comparing girls (boys) at single-sex schools with girls (boys) at coeducational schools. By gradually adding control variables, we check whether differences in student characteristics, family background (see Table 3.1), school characteristics (see Table A3.1), and teacher characteristics (see Table A3.2) alter the estimates. Since the random assignment process ensures that background characteristics are not correlated to a student's type of school, adding information on students' family backgrounds should not fundamentally alter the estimates. In contrast, differences in school resources and teacher characteristics are likely to be endogenous and might be related to the effect of single-sex schooling. For example, teachers are not randomly allocated to schools and it is possible that different types of teacher select into coeducational and single-sex schools. In addition, despite the fact that the Equalization Policy aimed at the equalization of schools and the ultimate reduction of differences in school quality, there are some differences in observable school characteristics between single-sex and coeducational schools (see Table A3.1 in the Appendix). Thus, by gradually controlling for school and teacher characteristics, it can be revealed if the effect of single-sex schooling operates through these channels.

Public and private schools in Korea are very similar and features such as selection of students, tuition fees, teacher salaries, and curriculum, are actually regulated (Kim and Lee, 2003). However, private schools maintained certain rights over personnel decisions and school facilities. Thus, the reported differences in the computer-student ratio and the level of autonomy in hiring might be driven by differences in the composition of private and public schools within coeducational and single-sex schools. Since teachers at public schools are required to rotate schools in 5 to 6 year intervals and we do not directly observe if a school is publicly or privately managed, we use the variable "Share of teachers

[17] Similarly, the fact that high-demanding parents are able to move to better-off school districts should not bias the estimate if the distribution of single-sex and coeducational schools is not related to characteristics of school districts. After moving, students would again be subject to the random allocation procedure across schools within the new district.

employed at a school for more that 5 years" (Share teacher > 5 years) as a proxy for private management in the regressions (Kim et al., 2008).[18]

In addition to differences in school and teacher characteristics across single-sex and coeducational schools, the dataset used in this study allows us to focus on more detailed information on the atmosphere and organization of schools. In particular, we account for differences in the disciplinary classroom and school climate, teaching practices, and student attitudes, all of which are often argued to be influential determinants in the public debate, to investigate the underlying mechanisms (see Section 3.6). Further, we check whether effects of single-sex education are hheterogeneous across student groups and divide the sample into students from different family backgrounds. In addition, quantile regressions reveal if the effects are hheterogeneous with respect to different parts of the performance distribution. Finally, we compare students across gender to see how girls at single-sex and coeducational schools actually perform relative to their male peers (see Section 3.7).

3.5 Results

3.5.1 The Effect of Single-Sex Schooling

Table 3.3 reports OLS estimates of the effect of single-sex schooling for girls and boys at middle schools.[19] We start with a univariate model with *attending a single-sex school* as the explanatory variable. The model is then – step by step – extended by controls on student and background characteristics, school characteristics, and teacher characteristics.[20] In contrast to student and background characteristics, which are pre-determined as students enter middle school, information on schools and teachers are potential channels through which single-sex schooling might affect student achievement.

The variable of interest, attending a single-sex school, is positive and significant in math for girls at middle schools throughout all specifications. Neither adding individual control variables (column 2) nor school (column 3) and teacher variables (column 4) alters

[18] Given the requirement to rotate schools after at most 6 years at a particular school, public schools should have a very low share of teachers that they employ for more than 5 years. Table A3.1 shows that for about 70 percent of coeducational and 71 percent of single-sex schools the share of teachers which are employed at the school for more than 5 years is at most 10 percent. This implies that the share of schools that are publicly managed is approximately equal across single-sex and coeducational schools.

[19] See Section 3.A in the Appendix for the results for high school students.

[20] See Table A3.3 in the Appendix for the complete model.

magnitude and significance of the coefficient of interest. Comparing columns (1) and (2) shows that the coefficient is robust to the inclusion of individual background variables. Given that observable student characteristics are likely to be correlated with unobserved student characteristics that determine student achievement, this encourages the random allocation of students to schools. The fact that school and teacher variables do not change the estimates indicates that conventional school and teacher characteristics are not driving the effect. Overall, these results suggest that girls at single-sex schools outperform girls at coeducational schools by about 12.4 percent of a standard deviation (column 4). For science, the coefficients are positive, but smaller and therefore not significant at conventional levels throughout the specifications.

The lower part of Table 3.3 reports the results for boys. We find insignificant coefficients, which are mainly close to zero for all specifications and both subjects. This indicates that there are neither beneficial nor adverse effects of single-sex schooling for boys at middle schools.

3.5.2 The Effect of Single-Sex Classes

Interestingly, there exist also single-sex classes at coeducational schools. In the sample of 53 coeducational schools, 15 classes are actually single-sex, which comes to seven all-boy classes and eight all-girl classes. In the analysis presented above, the effect of single-sex schooling is estimated by comparing students at single-sex schools to students at coeducational schools regardless whether they attend single-sex or mixed classes. However, many arguments for and against single-sex education are actually related to differences in the atmosphere between single-sex and mixed classes. Having students that attend single-sex classes at coeducational schools in the comparison group might therefore add noise to the analysis.

In the following, the effect of single-sex schooling is estimated in a sample that excludes single-sex classes at coeducational schools. The share of boys then ranges from 34 percent to 65 percent in mixed classes at coeducational schools. In addition, the effect of single-sex *classes* is estimated by comparing mixed classes to single-sex classes, regardless whether they are located within single-sex or coeducational schools. However, in contrast to the allocation of students to middle schools, the allocation of students across classes might be

Table 3.3
The Effect of Single-Sex Schooling on Student Achievement

	Female							
	Math				Science			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Single-Sex Schooling	0.108*	0.133***	0.111**	0.124**	0.060	0.076	0.073	0.078
	(0.059)	(0.050)	(0.051)	(0.048)	(0.057)	(0.053)	(0.054)	(0.054)
Student Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes
Imputation Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	2348	2348	2348	2348	2348	2348	2348	2348
Clusters	76	76	76	76	76	76	76	76
R^2	0.003	0.173	0.182	0.189	0.001	0.137	0.148	0.157
	Male							
	Math				Science			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Single-Sex Schooling	-0.096	-0.023	-0.010	-0.001	-0.091	-0.026	-0.026	-0.035
	(0.065)	(0.050)	(0.049)	(0.056)	(0.068)	(0.054)	(0.052)	(0.061)
Student Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes
Imputation Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	2427	2427	2427	2427	2427	2427	2427	2427
Clusters	78	78	78	78	78	78	78	78
R^2	0.002	0.176	0.186	0.200	0.002	0.167	0.175	0.183

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home for the teacher reported first if there are several. The regressions control for the fact that some students have several teachers in math and science. The complete model is reported in Table A3.3 in the Appendix. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

endogenous. Single-sex classes are therefore more likely to consist of a selected group of students, e.g., especially bad behaving or poor performing students.

Column (1) in Table 3.4 shows the effect of single-sex schooling for girls in math in a sample that is restricted to coeducational schools with mixed classes only. Compared to Table 3.3, the coefficient is slightly increased which translates into a higher significance. The effect for single-sex classes is almost identical (column 3). This suggests that the positive effect for girls is driven by differences between single-sex and mixed classrooms. For boys, the effect of single-sex schooling is larger than in Table 3.3 in the restricted sample, but statistically not significant. The coefficient on single-sex classes increases further in magnitude, pointing to beneficial effects of single-sex classes at coeducational schools for boys. However, since the number of single-sex classes at coeducational schools is very limited and the composition is potentially endogenous, these results should be taken with caution.

Overall, we find a positive effect of single-sex education on girls' performance which is not driven by differences in school and teacher characteristics between single-sex and coeducational schools. The effect is slightly more pronounced if single-sex classes at coeducational schools are either excluded from the analysis or are also classified as single-sex. Furthermore, the finding is restricted to girls' achievement in math. One explanation for this might be that math achievement is a better indicator of teacher instruction in class, compared to science, which is often not taught as a single subject. Thus, if more studious classrooms allow teachers in all-girl classes to cover the curriculum more extensively, this might be reflected in the large, significant coefficients for math achievement. Moreover, math is a traditionally male subject and the positive effects might be driven by less gender-stereotyped attitudes at single-sex schools.

In the following, these potential channels are investigated. Since the allocation of students to schools, in contrast to the allocation to classes, is random, we stay with the classification of single-sex and coeducational *schools*.

3.6 Channels

Given the positive and large effects for girls in math, it is important to understand the underlying mechanisms. One explanation for the positive effect for girls in math are

Table 3.4
Accounting for Single-Sex Classes at Coeducational Schools

	Excluding Single-Sex Classes at Coeducational Schools		Including Single-Sex Classes at Coeducational Schools	
	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
Single-Sex School	0.138***	0.057		
	(0.050)	(0.062)		
Single-Sex Class			0.131**	0.084
			(0.050)	(0.054)
Student Controls	Yes	Yes	Yes	Yes
School Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes
Observations	1967	2137	2348	2427
Clusters	68	71	76	78
R^2	0.187	0.191	0.189	0.201

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home for the teacher reported first if there are several. The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

differences in the peer composition between single-sex and coeducational schools. This is at the same time a robustness check, since the random assignment should be reflected in similar peer characteristics across schools. Furthermore, single-sex and coeducational schools might differ in the atmosphere and organization within schools. Even though some of these dimensions are unobservable, we are able to compare coeducational and single-sex schools in three influential areas – teaching practices, student attitudes toward math, including self-perceived competence, and disciplinary climate – which allows us to check whether the positive effect of single-sex schooling for girls can be explained by differences between single-sex and coeducational schools.

3.6.1 Peer Characteristics

Given that we are interested in the effect of single-sex education, it is important to separate gender compositional effects from other peer effects, such as advantageous family backgrounds and environments. Even though students in Korea are randomly assigned to schools within their residential districts, it is possible, but rather unlikely, that we observe only all-girl schools in better-off areas, whereas we observe coeducational schools in disadvantaged areas. If girls with advantageous family backgrounds are grouped within all-girl schools, the effects we find are not due to the absence of boys, but could be attributed to a better student composition at single-sex schools. For example, Jimenez and Lockheed (1989) attribute positive effects of single-sex schooling for girls in Thailand to favorable peer characteristics.

To make sure that it is the absence of boys in contrast to advantageous family background characteristics of female peers that drive the positive effects for girls in math, we account for the quality of a student's peers by including class-level measures of students' family backgrounds in the regression. This approach is especially comprehensive, since TIMSS assesses complete classes. In particular, we control for the share of peers with a low socio-economic background as measured by the number of books at home, the share of peers with high home resources as a proxy for wealth, the average family size, the share of peers with at least one parent holding a university degree, the share of peers with mothers holding a university degree, and the average amount of time spent studying by an individual's peers as a proxy for peer pressure. Table 3.5 reveals that the coefficient on the variable of interest is not affected by including those measures into the regression. Overall, Table 3.5 suggests that the effects are not due to selection.

3.6.2 Teaching Style

One argument made in favor of single-sex schooling is that such schools offer the opportunity to tailor schooling to each sex's unique needs. Differences in the way students are taught, therefore, might account for the positive effects found for girls. On the one hand, supporters of single-sex education claim that brain differences between boys and girls require different teaching styles.[21] On the other hand, less disruption and more studious classroom climates at all-girl schools might motivate teachers to give more homework or work more often in groups.[22] Table A3.4 in the Appendix shows that students at both all-girl and all-boy

[21] Neuroscientists have found only few brain differences between men and women and none of them have been linked to teaching practices.

[22] The fact that achievement gains can be driven by differences in teaching styles has been documented by Jürges and Schneider (2010), who attribute positive effects of central exit exams to the fact that students were required to work harder.

Table 3.5
The Effect of Single-Sex Schooling accounting for Peer Quality

	Female Math Test Score					
	Socio-economic Background			Parental University Education		Peer Pressure
Controlling for Share of Peers with	Few books	High home resources	Family size	At least one parent	Mother	Study time
	(1)	(2)	(3)	(4)	(5)	(6)
Single-Sex Schooling	0.134***	0.134**	0.130***	0.140***	0.155***	0.142***
	(0.049)	(0.052)	(0.049)	(0.051)	(0.052)	(0.044)
Aggregate Peer Measure	-0.211	0.161	-0.084	0.160	0.500*	0.733***
	(0.220)	(0.224)	(0.062)	(0.169)	(0.275)	(0.272)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2348	2346	2336	2348	2348	2281
Clusters	76	76	76	76	76	76
R^2	0.189	0.188	0.185	0.189	0.190	0.190
	Male Math Test Score					
Controlling for Share of Peers with	Few books	High home resources	Family size	At least one parent	Mother	Study time
	(1)	(2)	(3)	(4)	(5)	(6)
Single-Sex Schooling	0.020	0.015	-0.022	0.012	0.008	0.024
	(0.054)	(0.057)	(0.055)	(0.055)	(0.054)	(0.057)
Aggregate Peer Measure	-0.447***	0.223	-0.065	0.212	0.363**	0.486
	(0.152)	(0.174)	(0.059)	(0.140)	(0.168)	(0.369)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2427	2425	2397	2427	2427	2349
Clusters	78	78	78	78	78	78
R^2	0.202	0.201	0.201	0.200	0.201	0.196

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. "Aggregate Peer Measure" refers to the share of a student's peers that belongs to the respective category. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

schools report more often that they "copy notes from the board" compared to students at coeducational schools. Table A3.4 also suggests that students from all-girl schools get homework more frequently and that they work more often in groups. However, the reported differences are quite small.

Table 3.6 reports the effect of single-sex schooling while controlling for these measures of teaching style. The frequency of "having tests" and "giving homework" seems to be positively associated with student learning, however, most of the other measures are insignificant. Most importantly, the coefficient on the variable single-sex schooling does not change in magnitude or significance for either girls or boys. This suggests that the effects of single-sex schooling are not driven by differences in the observed dimensions of teaching practices.

3.6.3 Student Attitude

Student attitudes toward math present another possible channel and the one most closely related to the literature on gender stereotypes. The construction of gender identities at schools is especially important with regard to the persisting gender test score gap in math (see, e.g., Guiso et al., 2008; Fryer and Levitt, 2010) and the low representation of women in STEM-related occupations (Organisation for Economic Co-operation and Development, 2008b).[23] The presence of the opposite sex at mixed schools may either deforce or reinforce gender-stereotyped attitudes and thereby influence the likelihood that boys (girls) engage in stereotypically female (male) subjects or fields (e.g., Thompson, 2003; Joshi et al., 2010; Favara, 2011; Halpern et al., 2011).[24] If single-sex education leads to gender-atypical educational choices and increases girls' interest in math, this is likely to improve learning and could therefore explain the positive effects found for girls. Furthermore, the presence of boys in the classroom could be especially intimidating for girls in a stereotypically male subject such as math. Given a predominant opinion that boys outperform girls in math, a girl at a coeducational school is more likely to assess herself poorly relative to her peers, which include girls *and* boys, compared to a girl at a single-sex school.[25]

[23] The role of gender identities is based on Akerlof and Kranton (2000).
[24] Favara (2011) confirms that subject choices of girls at single-sex schools are more similar to those of their male schoolmates. In contrast, Halpern et al. (2011) show that sex segregation increases gender stereotyping.
[25] Beyer and Bowden (1997) show that females' self-perceptions of performance were inaccurately low in "male" tasks.

Table 3.6
The Effect of Single-Sex Schooling accounting for Teaching Practices

	Female Math Test Score			
	Copying Notes	Having a Test/ Quiz	Giving Homework	Working in Groups
	(1)	(2)	(3)	(4)
Single-Sex Schooling	0.123**	0.141***	0.121**	0.127**
	(0.048)	(0.050)	(0.049)	(0.049)
Once in a while	0.096	0.212***	0.195**	-0.088**
	(0.092)	(0.064)	(0.086)	(0.043)
Pretty often/ Always	0.101	0.222***	0.228**	0.053
	(0.088)	(0.082)	(0.088)	(0.062)
Full Controls	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes
Observations	2348	2348	2348	2348
Clusters	76	76	76	76
R^2	0.189	0.195	0.191	0.191

	Male Math Test Score			
	Copying Notes	Having a Test/ Quiz	Giving Homework	Working in Groups
	(1)	(2)	(3)	(4)
Single-Sex Schooling	0.014	0.012	0.003	0.004
	(0.056)	(0.054)	(0.055)	(0.056)
Once in a while	0.030	0.404***	0.192**	-0.017
	(0.090)	(0.059)	(0.082)	(0.044)
Pretty often/ Always	-0.047	0.362***	0.101	0.037
	(0.095)	(0.067)	(0.088)	(0.064)
Full Controls	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes
Observations	2427	2427	2427	2427
Clusters	78	78	78	78
R^2	0.198	0.218	0.200	0.197

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A3.5 reports descriptive statistics on several indicators of student attitudes toward math, their self-perceived competence in math, and their educational aspirations. Girls at coeducational schools seem to have a less positive attitude toward math compared to girls at single-sex schools; apart from that, however, there are only very small differences regarding their confidence, educational aspirations, and preferences for math. Nevertheless, we control for the reported measures of student attitudes in the regressions and report the corresponding estimates in Table 3.7. If the coefficient of interest, single-sex schooling, is capturing some of these differences, the coefficient should decrease in size and significance. As expected, all the measures of student attitudes are positively and significantly associated with better math results for both, boys and girls. However, although the measures of student attitude have a strong explanatory power for student achievement, the positive effect of single-sex schooling remains significant and is not affected by differences in the attitude toward math.

3.6.4 Disciplinary Climate

The most obvious reason why single-sex education might be especially beneficial for girls involves the relatively more restless and disruptive behavior of boys. There exists evidence that a larger share of girls in a class is associated with higher academic achievement which can partly be explained by a lower level of classroom disruption and violence (Hoxby, 2000; Lavy and Schlosser, 2011). TIMSS reports students', teachers', and principals' perceptions on several aspects of the disciplinary climate of classrooms and schools. Table A3.6 in the Appendix reveals that, according to teacher and principal reports, there are indeed differences in the disciplinary climate at coeducational, all-girl, and all-boy schools.

Teachers are asked to what extent teaching is hindered by disruptive students, uninterested students, a wide range of backgrounds, and a wide range of academic abilities. Twenty-six to 29 percent of students at coeducational schools attend classrooms where "disruptive" and "uninterested" students are reported to be "a serious problem". These fractions are somewhat smaller for all-boy schools and about half the size at all-girl schools. Further, teachers at coeducational schools perceive "a wide range of backgrounds" and "a wide range of academic abilities" more often as a problem compared to teachers at single-sex schools. Moreover, at almost 70 percent of all-girl schools, the "injury of students" is "not a problem at all", but there are large fractions of coeducational and all-boy schools

Table 3.7

The Effect of Single-Sex Schooling accounting for Student Attitude

	Female Math Test Score				
	Like Math	Positive Attitude	Important for myself	Confidence in Math	Educational Aspirations
	(1)	(2)	(3)	(4)	(5)
Single-Sex Schooling	0.122***	0.102**	0.101**	0.132***	0.110**
	(0.044)	(0.046)	(0.047)	(0.045)	(0.044)
Medium	0.597***	0.364***	0.296***	0.720***	0.502***
	(0.040)	(0.045)	(0.081)	(0.077)	(0.074)
High	0.974***	0.926***	0.528***	1.375***	1.030***
	(0.049)	(0.060)	(0.081)	(0.096)	(0.054)
Full Controls	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes
Observations	2342	2339	2341	2342	2342
Clusters	76	76	76	76	76
R^2	0.305	0.248	0.211	0.251	0.286
	Male Math Test Score				
	(1)	(2)	(3)	(4)	(5)
Single-Sex Schooling	0.012	0.011	0.013	0.008	0.007
	(0.046)	(0.053)	(0.052)	(0.050)	(0.050)
Medium	0.614***	0.320***	0.387***	0.583***	0.409***
	(0.037)	(0.042)	(0.063)	(0.088)	(0.084)
High	0.838***	0.729***	0.534***	1.163***	0.897***
	(0.053)	(0.072)	(0.062)	(0.092)	(0.058)
Full Controls	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes
Observations	2414	2413	2416	2412	2416
Clusters	78	78	78	78	78
R^2	0.304	0.237	0.220	0.245	0.281

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

that report it at least as “a minor problem”. This indicates that the disciplinary climate, as reported by both teachers and principals, is rougher at coeducational schools, which might be especially detrimental to girls’ achievement.

Table 3.8 reports the effect of single-sex schooling while accounting for the different measures of disciplinary climate. While there is a positive association between a “good” disciplinary climate and student achievement, the coefficient of interest is not influenced by the measures reported by students (columns 1 and 2). Furthermore, the effect of single-sex schooling is unchanged if the regression accounts for the extent of “disruptive students” and “intimidation of students” at school. Given that there are no differences in the descriptive statistics regarding these measures, this is actually not surprising.

In contrast, Table 3.8 reveals a negative and strong association between teachers who report “uninterested students as a great problem” and girls’ achievement. The coefficient of single-sex schooling drops by one quarter and loses significance (see column 4 in the upper part of Table 3.8).[26] Similarly, the “injury of students” presents a larger problem at coeducational schools and reduces the estimate of single-sex schooling considerably for girls. Although some teachers report “differences in students’ backgrounds as a problem”, this is not reflected in lower achievement by students. However, teachers at coeducational schools report more often that “differences in students’ math abilities” limit their teaching. This is also reflected in lower student achievement and reduces the estimate of single-sex schooling. Since students are randomly assigned to single-sex and coeducational schools, the variation in math ability at each type of school should initially be similar. One explanation for this finding might be that achievement of students in general, or of boys and girls in particular, at coeducational schools has diverged over time. Alternatively, it might be that teachers of coeducational classes just perceive abilities as more diverse, possibly due to a predetermined opinion that boys outperform girls in math, which is then correlated with lower achievement of girls.

Overall, these results suggest that differences in teaching practices and student attitudes cannot explain the achievement gains for girls at single-sex schools. However, Table 3.8 indicates that the positive effects can partly be explained by a rougher classroom atmosphere at coeducational schools.

[26] There are a few missings in the reported measures by teachers and principals which results in a smaller number of observations. We checked, however, that the decrease in the size of the coefficient is not driven by the smaller sample size.

Table 3.8

The Effect of Single-Sex Schooling accounting for Disciplinary Climate

	Female Math Test Score							
	Students' Behavior		Students are		Wide Range of		Students get	
	Orderly	As told	Disruptive	Uninterested	Backgrounds	Acad. Abilities	Intimidated	Injured
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Single-Sex	0.110**	0.123**	0.114**	0.082	0.122**	0.074	0.105**	0.070
	(0.049)	(0.048)	(0.050)	(0.051)	(0.048)	(0.046)	(0.050)	(0.046)
Medium	0.018	0.215**	-0.007	-0.052	0.057	-0.205***	-0.020	-0.144***
	(0.077)	(0.098)	(0.063)	(0.065)	(0.066)	(0.074)	(0.065)	(0.051)
High	0.192**	0.210**	0.024	-0.180*	0.010	-0.089	-0.182**	-0.095
	(0.091)	(0.099)	(0.083)	(0.096)	(0.110)	(0.075)	(0.075)	(0.074)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2348	2348	2281	2281	2281	2327	2249	2249
Clusters	76	76	74	74	74	75	73	73
R^2	0.195	0.191	0.192	0.194	0.192	0.193	0.190	0.190
	Male Math Test Score							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Single-Sex	0.001	0.003	-0.030	-0.015	0.011	-0.017	-0.002	-0.010
	(0.055)	(0.056)	(0.060)	(0.064)	(0.053)	(0.060)	(0.056)	(0.054)
Medium	0.126**	0.149**	0.018	-0.049	0.106**	0.005	-0.017	0.016
	(0.062)	(0.064)	(0.067)	(0.079)	(0.053)	(0.061)	(0.053)	(0.053)
High	0.104	0.085	-0.110	-0.078	0.081	-0.096	-0.035	0.116
	(0.069)	(0.071)	(0.072)	(0.074)	(0.071)	(0.117)	(0.096)	(0.083)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2427	2427	2401	2401	2401	2401	2411	2411
Clusters	78	78	77	77	77	77	77	77
R^2	0.199	0.199	0.200	0.199	0.200	0.199	0.197	0.197

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

3.7 Heterogeneous Effects

In a next step, we investigate whether the effect varies across different family backgrounds and across levels of the performance distribution. Finally, we investigate differences in achievement across gender.

3.7.1 Family Background

Paying attention to students with relatively less supportive backgrounds is important for several reasons. First, it has been argued that either type of schooling might be more beneficial or harmful to some students than to others. For example, Riordan (1990) shows that the greatest gains in single-sex schooling are those experienced by Hispanic and African-American males and females at schools with large minority populations. One reason for this might be that students with low socio-economic backgrounds typically receive less support at home in studying and, since their education depends more strongly on instruction received at school, respond more strongly to it. Another reason might be that students belonging to minorities – either ethnic or socio-economic – are easily intimidated and need a great deal of attention or support. Paying attention to students from a low socio-economic background is also politically relevant since those students are at a higher risk of dropping out of school or performing very poorly, which might come at a high cost for the society as a whole (see, e.g., Organisation for Economic Co-operation and Development, 2009; Hanushek and Woessmann, 2011a).

The number of books at home has often been used as a measure for socio-economic background in the literature (see, e.g., Woessmann, 2003, 2008; Schütz et al., 2008) and is a strong predictor of academic achievement for both girls and boys (see Table A3.3 in the Appendix). Thus, we divide the sample into two groups of equal sizes and classify students with less than 100 books at home as students with relatively low socio-economic backgrounds and those with more than 100 books at home as students with relatively high socio-economic backgrounds. We further generate a variable that takes the value 1 if students have relatively low educated parents since parental education is a strong indicator of parents' interest in their children's educational aspirations and developments. Similarly, a variable indicating whether a student reports that his or her mother is not interested in his or her math achievement is generated. We then interact those measures with the variable of single-sex schooling and include them separately in the regression.

As expected, Table 3.9 shows that a low socio-economic background, low educated parents, as well as uninterested mothers are strongly and negatively associated with girls' and boys' math achievements. Interestingly, the interaction of all three measures of a non-supportive background and attending a single-sex school are large in magnitude and significant for girls. In particular, positive effects of single-sex schooling are even restricted to girls with low socio-economic or low educated family backgrounds, while the effects are larger for girls who report that their mothers are not interested in their math achievements. Consistent with the previous finding for boys, the effect of single-sex schooling remains insignificant and around zero and the coefficients on the interactions are not significant.

3.7.2 Performance Distribution

Table 3.9 reveals that the positive effect of single-sex schooling is restricted to girls with low socio-economic backgrounds. In the following, we check whether beneficial effects are especially large for low performing or high performing girls.

Table 3.10 reports quantile regressions for girls with low socio-economic backgrounds and girls from high socio-economic backgrounds. While a positive, significant effect of single-sex schooling can be reported for the 25., the 50., and the 75. percentile, the coefficients are largest for students in the lower part of the test score distribution. In contrast, the lower part of Table 3.10 reveals that there are no beneficial effects for students from high socio-economic backgrounds, regardless whether they perform relatively poor or high.

3.7.3 Test Score Gender Gap

So far, the analysis compared girls (boys) at coeducational schools with girls (boys) at single-sex schools. The results suggest that girls from low parental support backgrounds at single-sex schools outperform girls from low parental support backgrounds at coeducational schools, whereas there are no significant differences for boys. However, given the existence of a gender test score gap in math, it is also interesting how girls at single-sex and coeducational schools perform relative to boys.

Table 3.9
Heterogeneous Effects by Family Background

	Test Score Math					
	Female			Male		
	(1)	(2)	(3)	(4)	(5)	(6)
Single-Sex Schooling	0.035	0.033	0.086*	-0.032	-0.036	0.015
	(0.056)	(0.046)	(0.048)	(0.069)	(0.064)	(0.052)
Low Socio-economic	-0.557***			-0.441***		
	(0.058)			(0.054)		
Low x Single-Sex	0.190**			0.033		
	(0.072)			(0.079)		
Low-educated		-0.373***			-0.331***	
		(0.055)			(0.059)	
Low-educated x Single-Sex		0.236***			0.085	
		(0.070)			(0.084)	
Math not important for Mother			-0.711***			-0.782***
			(0.177)			(0.134)
NotImp x Single-Sex			0.552**			0.074
			(0.220)			(0.180)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2348	2348	2335	2427	2427	2406
Clusters	76	76	76	78	78	78
R^2	0.162	0.163	0.195	0.162	0.176	0.224

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Low (high) socio-economic background refers to students with less (more) than 100 books at home. Low-educated refers to students with parents with less than secondary education. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.10
Heterogeneous Effects by Test Score Distribution

	Female Math Test Score								
	Low Socio-economic Background								
	.25 Quantile			.50 Quantile			.75 Quantile		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Single-Sex Schooling	0.252***	0.232***	0.313***	0.164**	0.171**	0.139*	0.174***	0.177**	0.200**
	(0.090)	(0.074)	(0.076)	(0.068)	(0.069)	(0.080)	(0.064)	(0.085)	(0.101)
Student Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School Controls	No	No	Yes	No	No	Yes	No	No	Yes
Teacher Controls	No	No	Yes	No	No	Yes	No	No	Yes
Imputation Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	1264	1264	1264	1264	1264	1264	1264	1264	1264
Pseudo R^2	0.0080	0.0793	0.0991	0.0033	0.0568	0.0701	0.0049	0.0346	0.0514
	High Socio-economic Background								
	.25 Quantile			.50 Quantile			.75 Quantile		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Single-Sex Schooling	0.040	0.042	0.039	-0.029	0.034	-0.025	-0.033	0.071	0.064
	(0.094)	(0.080)	(0.095)	(0.060)	(0.064)	(0.102)	(0.076)	(0.059)	(0.090)
Student Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School Controls	No	No	Yes	No	No	Yes	No	No	Yes
Teacher Controls	No	No	Yes	No	No	Yes	No	No	Yes
Imputation Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	1084	1084	1084	1084	1084	1084	1084	1084	1084
Pseudo R^2	0.0001	0.0592	0.0913	0.0002	0.0401	0.0652	0.0002	0.0502	0.0749

Notes: Data source: TIMSS 1999. This table reports quantile regressions. Bootstrapped standard errors with 100 replications are reported in parentheses. Low (high) socio-economic background refers to students with less (more) than 100 books at home. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher's age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.11 shows a pooled regression divided by socio-economic background. The coefficient on the female dummy can be interpreted as the gender test score gap in math. Without controlling for single-sex schools, there is no significant difference in math achievement between boys and girls in the full sample (column 1). However, as soon as the regression controls for all-boy (coefficient on single-sex) and all-girl schools (coefficient on the interaction of single-sex and female), the coefficient on the female dummy turns negative and significant, revealing the well-known gender test score gap in math. In other words, columns (1) and (2) show that there are no significant differences in math achievement between boys at either coeducational schools or single-sex schools and girls at single-sex schools. However, girls at coeducational schools underperform boys at both school types and girls at single-sex schools.

Table 3.11 also reports the results for students with low and high socio-economic backgrounds as measured by books at home (columns 4 to 9). Interestingly, there is no test score gap in math between boys and girls from relatively high socio-economic backgrounds (column 7), not even after controlling for single-sex schools (column 8). In contrast, the test score gap between boys and girls with low socio-economic backgrounds at coeducational schools is especially large (column 5), but decreases strongly as soon as the regression additionally controls for "the extent of injuries" (column 6), – which can be viewed as a proxy for disciplinary climate at schools. Altogether, this in-depth analysis suggests that girls from less supportive backgrounds fall behind at coeducational schools and that the atmosphere at coeducational schools plays an important role in this result.

3.8 Robustness Tests

The results suggest that single-sex schooling is beneficial for girls from low parental support backgrounds in math, but does not have any effects for boys. The causal interpretation in an ordinary least squares approach is based on the assumption that attendance at single-sex schools is orthogonal to student characteristics such as socio-economic background and ability. Since students in Korea are randomly assigned to schools, this is very plausible. Tables 3.1 and 3.2 lend support to this assumption by reporting very small and mostly non-significant differences in a very rich set of observable student characteristics. The few significant differences in student characteristics may be driven by differences in the location

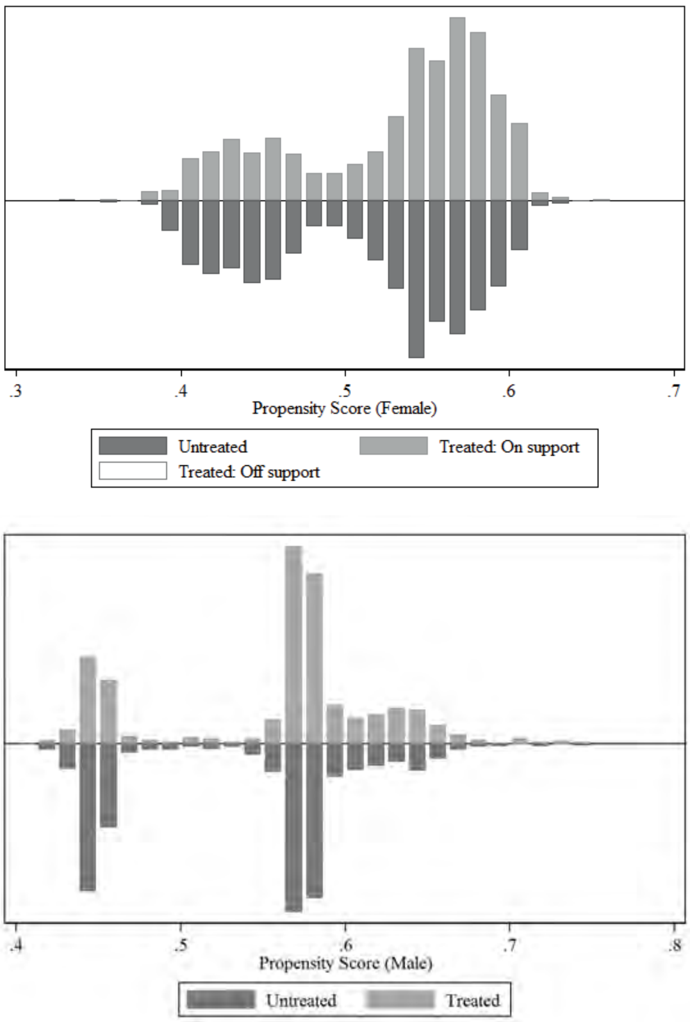
Table 3.11
The Effect of Single-Sex Schooling and the Test Score Gender Gap in Math

	Math								
	Socio-economic Background								
	All			Low			High		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.040	-0.092**	-0.034	-0.071	-0.170***	-0.083	0.012	0.007	0.048
	(0.034)	(0.040)	(0.052)	(0.046)	(0.051)	(0.070)	(0.038)	(0.050)	(0.064)
Single-Sex Schooling		0.037	0.027		0.053	0.044		0.024	0.014
		(0.047)	(0.052)		(0.061)	(0.068)		(0.067)	(0.072)
Single-Sex x Female		0.090	0.072		0.164*	0.139		0.008	-0.008
		(0.067)	(0.071)		(0.087)	(0.090)		(0.083)	(0.089)
Injured2			0.036			0.021			0.053
			(0.049)			(0.062)			(0.069)
Injured3			0.047			0.074			0.057
			(0.062)			(0.091)			(0.082)
Injured2 x Female			-0.126*			-0.166*			-0.097
			(0.067)			(0.089)			(0.086)
Injured3 x Female			-0.067			-0.170			-0.010
			(0.073)			(0.106)			(0.111)
Full Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Imputation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4775	4775	4660	2575	2575	2500	2200	2200	2160
Clusters	116	116	113	116	116	113	116	116	113
R^2	0.181	0.183	0.182	0.124	0.129	0.130	0.110	0.110	0.112

Notes: Data source: TIMSS 1999. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Low (high) socio-economic background refers to students with less (more) than 100 books at home. “Injured” refers to the extent of injuries at a school reported by the principal. Student controls include age, parents’ education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and student-computer-ratios, share of teachers employed for longer than 5 years, and hiring and course autonomy. Teacher controls include teacher’s age, gender, education and books at home (for the teacher reported first if there are several). The regressions control for the fact that some students have several teachers in math and science. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

of schools, for which we cannot control in the analyses. However, Table 3.5 shows that controlling for peer quality as a proxy for neighborhood characteristics does not change the estimates. Similarly, controlling for the extensive set of background variables reported in Table 3.1 and 3.2 at the individual and the class level leaves the estimate unchanged (see columns 8 and 11 in Table 3.12).

Figure 3.2. Estimated Propensity Scores by Gender.



Notes: The figure reports estimated propensity scores by gender after performing kernel matching using students' age, parent's education, books at home, and living with both mother and father as explanatory variables. Data source: TIMSS 1999.

In addition, we perform propensity score matching analyses and compare students at single-sex and coeducational schools who have similar estimated propensities to attend single-sex schools based on their observable characteristics, namely age, parent's education, books at home, and living with both mother and father (see Figure 3.2). We perform three common matching techniques – namely kernel, nearest-neighbor, and radius matching.[27] We impose a common support which means that we drop treated individuals that have an estimated propensity score less than the minimum or more than the maximum of the controls. While all three techniques adjust for pre-treatment observable differences, they use different procedures and weighting schemes. The kernel matching technique lends a higher weight to non-treated observations that are more similar in terms of the estimated propensity score. For the nearest neighbor matching, the non-treated observation for each treated observation that is closest in terms of the estimated propensity score is chosen as matching partner. The radius matching uses only non-treated observations with estimated propensity scores that lie within a specified radius.

Table 3.12 reports the results from the propensity score analysis. As expected given the small and few differences in student characteristics and the large similarity in estimated propensity scores (see Figure 3.2), the point estimates are quite similar to the OLS estimates for both the conventional student background control set (reported in Table 3.1) and the extensive student background control set (reported in Tables 3.1 and 3.2). Overall, we find positive, significant effects for girls at single-sex schools, but no effects for boys in either OLS or propensity score analysis.

3.9 Conclusions

This paper contributes to a growing quasi-experimental literature on single-sex education and investigates the effect of single-sex schooling in a particularly interesting setting. In the Korean education system, students are randomly assigned to secondary schools, which can be either single-sex or coeducational. Given that attendance at single-sex schools is orthogonal to student characteristics such as socio-economic background and ability, the comparison between girls (boys) at coeducational schools and girls (boys) at single-sex schools should identify a reliable effect of single-sex schooling on student achievement. Moreover, the rich dataset used in this study allows us to investigate a large set of potential channels and features that are often associated with single-sex schooling in the public

[27] See Leuven and Sianesi (2003).

Table 3.12

Robustness: Matching Estimates and OLS Estimates with Extensive Set of Control

	Female										
	Propensity Score Matching						OLS Estimates controlling at				
	Kernel		Near. Neighbor		Radius		Individual level		Class level		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Single-Sex Schooling	0.103***	0.090**	0.087*	0.078*	0.105***	0.092**	0.124***	0.118***	0.118**	0.110*	0.144**
	(0.042)	(0.043)	(0.046)	(0.046)	(0.042)	(0.043)	(0.048)	(0.049)	(0.055)	(0.061)	(0.064)
Student Background Variables											
Set 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Set 2	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	2347	2247	2347	2247	2341	2247	2348	2260	2348	2348	2348
LR-chi2/ R^2	37.41	59.85	37.41	59.85	37.41	59.85	0.187	0.231	0.027	0.044	0.051
	Male										
	Propensity Score Matching						OLS Estimates controlling at				
	Kernel		Near. Neighbor		Radius		Individual level		Class level		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Single-Sex Schooling	−0.038	0.000	−0.021	−0.010	−0.032	−0.001	−0.001	0.016	-0.004	0.046	0.088
	(0.042)	(0.043)	(0.045)	(0.046)	(0.042)	(0.043)	(0.056)	(0.053)	(0.053)	(0.060)	(0.064)
Student Background Variables											
Set 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Set 2	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	2427	2310	2427	2310	2427	2310	2427	2310	2427	2427	2427
LR-chi2/ R^2	49.25	93.24	49.25	93.24	49.25	93.24	0.199	0.260	0.042	0.074	0.079

Notes: Data source: TIMSS 1999. This table reports propensity score matching results and ordinary least square results. Standard errors (clustered at the school level for OLS results) in parentheses. Student control set 1 includes age, parents' education, books at home and living with mother and father. Student control set 2 includes computer at home, internet at home, calculator at home, frequency of reading a book at home, frequency of watching news or documentaries at home, frequency of going to the movies, frequency of watching opera or ballet, and frequency of watching comedies. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

debate. Since the underlying mechanisms of the effects of single-sex schooling are yet not well understood, this is an important contribution to the literature.

We find substantial, positive and significant effects of single-sex schooling for girls from low parental support backgrounds in math, which are largest among the group of low-performing students. In contrast, there are neither beneficial nor adverse effects for boys. Differences in school and teacher characteristics, gender-tailored education practices, or reduced gender stereotypes at single-sex schools cannot explain this finding. However, part of the effect can be attributed to a rougher disciplinary climate at coeducational schools.

Moreover, the comparison of boys and girls reveals that the test score gender gap in math is especially large for students from low parental support backgrounds who attend coeducational schools. In contrast, girls from single-sex schools and boys from either single-sex or coeducational schools perform equally well. This result suggests that girls from low parental support backgrounds might be somehow harmed by the presence of boys when learning a stereotypically male subject such as math. Given that most Western countries report large gender test score gaps in math while educating their students in coeducational schools, this is a particular interesting finding (see also Guiso et al., 2008; Fryer and Levitt, 2010).

The identification of causal effects relies on the underlying assumption that student characteristics are orthogonal to attendance at single-sex schools. While the Korean setting and several robustness checks suggest that the effects are not driven by differences in the types of students who attend single-sex and coeducational schools, it is crucial to control for the location of schools in further studies. Regarding the external validity of the results, it must be remembered that schooling tradition and culture in Korea obviously differs from that of Western societies. In addition, the data we analyze relate to a point of time when gender equality levels, as measured by, for example, the gender gap index (GGI), were relatively low in Korea. Even though this raises concerns about the generalizability of the findings, this paper documents an interesting pattern that is consistent with earlier findings. The fact that the in-depth analysis cannot fully explain the positive effects for girls suggests that future research should focus more on within class interactions when trying to understand the underlying mechanisms of the effects of single-sex schooling.

3.A High School Students

We use the Programme for International Student Assessment (PISA) that provides data on the educational achievement of 15-year-old students, which refers to the tenth grade at general and vocational high schools. The dataset provide extensive background information at the student level as well as information on school characteristics. Table A3.7 reports descriptive statistics on student characteristics for both groups of students. To indicate the single-sex status of a school, we rely on information as to the number of girls and boys enrolled at a school.

As described above, the High School Equalization Policy became subject of critique and the implementation was slowed down. Metropolitan cities continued to be required to follow the policy and assign their students to general high schools, but it was optional for smaller cities and rural school districts. To reduce the threat of selection, we focus on high schools in metropolitan areas and large cities, because they are most likely to be targeted by the Equalization Policy. The resulting dataset totals 4,390 individual observations at general high schools.

Table A3.8 reports the effects of single-sex schooling for girls and boys at high schools in math, science, and reading. The results for boys are similar to the results for middle schools: the coefficients for science and reading are close to zero and none of them is significant. The coefficients for girls are mostly positive, however, the standard errors are quite large and the coefficients are not significant. After controlling for school characteristics, the coefficient in math for girls at general high schools is about the same size as for middle school students (see Table 3.3 column 3). However, the coefficient is statistically not significant. Since we cannot exclude the possibility that the sample includes students that actually chose their school, this adds noise to the analysis and the coefficients are estimated less precise. Overall, the results of single-sex schooling on high school students confirm to some extent the results on middle school students. However, given that there are a number of exceptions for the High School Equalization Policy, these results should be taken with caution.

Table A3.1
Descriptive Statistics: Middle Schools

	Middle Schools			
	Coed	Single-Sex Schools		
	All	All	Female	Male
Total Enrollment	1316.70	1177.95	1204.23	1154.06
	(478.24)	(340.33)	(312.96)	(361.88)
Outskirts of a City	0.52	0.34	0.34	0.35
	(0.50)	(0.47)	(0.47)	(0.48)
Center of a City	0.48	0.66	0.66	0.65
	(0.50)	(0.47)	(0.47)	(0.48)
Student-Teacher-Ratio	25.30	24.82	24.59	25.03
	(5.64)	(4.07)	(3.06)	(4.80)
Student-Computer-Ratio	38.88	62.54	50.19	73.76
	(37.71)	(153.48)	(54.95)	(204.87)
Share Teacher $>$ 5 years	13.91	23.20	27.00	19.74
	(25.87)	(34.49)	(38.21)	(30.32)
Share Teacher $>$ 5 years $\leq$10 Percent	0.71	0.70	0.66	0.73
	(0.45)	(0.46)	(0.47)	(0.44)
Hiring Autonomy	0.29	0.42	0.45	0.39
	(0.43)	(0.49)	(0.49)	(0.49)
Courses Autonomy	0.92	0.94	0.95	0.94
	(0.26)	(0.24)	(0.23)	(0.25)
Students with Low Socio-economic Background	0.53	0.56	0.55	0.57
	(0.50)	(0.50)	(0.50)	(0.50)
Observations	2195	2561	1226	1335

Notes: Data source: TIMSS 1999. Individual observations weighted by sampling probabilities. Standard deviations in parentheses. Students with less than 100 books at home are classified as students with low socio-economic background.

Table A3.2
Descriptive Statistics: Teacher at Middle Schools

	Middle Schools			
	Coed	Single-Sex Schools		
	All	All	Female	Male
Female	0.65	0.59	0.79	0.42
	(0.48)	(0.49)	(0.41)	(0.49)
Master/ Phd	0.16	0.11	0.09	0.12
	(0.37)	(0.31)	(0.29)	(0.32)
Age: Under 30	0.29	0.13	0.16	0.10
	(0.45)	(0.34)	(0.37)	(0.30)
Age: 30-50	0.62	0.69	0.78	0.61
	(0.48)	(0.46)	(0.42)	(0.49)
Age: > 50	0.09	0.18	0.06	0.29
	(0.29)	(0.38)	(0.24)	(0.45)
Teaching Experience	12.34	13.21	10.64	15.55
	(8.54)	(9.38)	(7.72)	(10.12)
Books at Home: Up to 1 bookcase	0.25	0.38	0.32	0.44
	(0.43)	(0.49)	(0.47)	(0.50)
Books at Home: 2 bookcases	0.28	0.30	0.33	0.26
	(0.45)	(0.46)	(0.47)	(0.44)
Books at Home: 3-4 bookcases	0.47	0.32	0.35	0.29
	(0.50)	(0.47)	(0.48)	(0.46)
Observations	2195	2561	1226	1335

Notes: Data source: TIMSS 1999. Individual observations weighted by sampling probabilities. Standard deviations in parentheses.

Table A3.3
The Effect of Single-Sex Schooling on Student Achievement: Complete Model

	Female		Male	
	Math	Science	Math	Science
	(1)	(2)	(3)	(4)
Single-Sex Schooling	0.124**	0.078	-0.001	-0.035
	(0.048)	(0.054)	(0.056)	(0.061)
Age	0.046	0.156***	0.010	0.075
	(0.048)	(0.053)	(0.052)	(0.059)
Parents' Education: None	0.310**	0.228	0.519***	0.572***
	(0.146)	(0.149)	(0.120)	(0.095)
Parents' Education: Primary	0.343***	0.377***	0.377***	0.315***
	(0.112)	(0.109)	(0.084)	(0.077)
Parents' Education: Secondary	0.382***	0.410***	0.444***	0.420***
	(0.097)	(0.098)	(0.074)	(0.071)
Parents' Education: University	0.758***	0.661***	0.695***	0.624***
	(0.100)	(0.100)	(0.076)	(0.076)
Books at Home: 11-25	0.348***	0.267***	0.242**	0.203**
	(0.106)	(0.089)	(0.096)	(0.083)
Books at Home: 26-100	0.609***	0.522***	0.604***	0.516***
	(0.092)	(0.090)	(0.067)	(0.051)
Books at Home: 101-200	0.861***	0.731***	0.759***	0.681***
	(0.096)	(0.094)	(0.071)	(0.061)
Books at Home: > 200	1.039***	1.012***	0.985***	0.934***
	(0.088)	(0.085)	(0.078)	(0.063)
Live with Parents	0.135*	-0.058	0.209***	0.094
	(0.071)	(0.065)	(0.063)	(0.064)
Total Enrollment	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Center of a City	0.082	0.102	0.080	0.161***
	(0.056)	(0.065)	(0.060)	(0.061)
Student-Teacher-Ratio	-0.015*	-0.007	-0.004	0.009
	(0.008)	(0.010)	(0.006)	(0.007)
Student-Computer-Ratio	-0.001**	0.000	-0.000	-0.000***
	(0.001)	(0.000)	(0.000)	(0.000)
Share Teacher > 5 years	-0.001	-0.002***	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Hiring Autonomy	0.108*	0.082	0.009	-0.007
	(0.063)	(0.051)	(0.047)	(0.058)
Course Autonomy	-0.045	0.122	0.257***	0.217**
	(0.104)	(0.112)	(0.083)	(0.108)
Female Teacher	-0.101	-0.061	0.013	0.019
	(0.074)	(0.089)	(0.066)	(0.061)

Continued on next page

– continued from previous page –

	Female		Male	
	Math	Science	Math	Science
Teacher's Age: 30-50	0.035	-0.030	-0.144	-0.091
	(0.082)	(0.086)	(0.094)	(0.085)
Teacher's Age: > 50	-0.354*	-0.362**	-0.442**	-0.235
	(0.209)	(0.159)	(0.187)	(0.165)
Master/ Phd	-0.030	0.060	0.087	0.033
	(0.101)	(0.069)	(0.111)	(0.096)
Teaching Experience	0.000	0.003	0.008	0.005
	(0.006)	(0.005)	(0.006)	(0.005)
Teacher's Books: Up to 1 bookcase	-0.055	-0.009	0.147*	0.162**
	(0.062)	(0.066)	(0.081)	(0.081)
Teacher's Books: 2 bookcases	-0.043	-0.174**	0.086	0.030
	(0.065)	(0.068)	(0.093)	(0.079)
Several Teacher	0.121	-0.033	-0.278*	-0.195
	(0.112)	(0.168)	(0.162)	(0.202)
Constant	-1.608**	-3.184***	-1.586**	-2.451***
	(0.724)	(0.781)	(0.778)	(0.857)
Imputation Controls	Yes	Yes	Yes	Yes
Observations	2348	2348	2427	2427
Clusters	76	76	78	78
R^2	0.189	0.157	0.200	0.183

Notes: Data source: TIMSS 1999. This table reports the complete model for the specification reported in Table 3.3. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Reference category for parents' education is "do not know", for number of books at home is "0-10 books", for school location is "outskirts of a city", for teacher's age is "below 30", for teacher's education is a "bachelor degree", and for teacher's books at home is "more than 2 bookcases". Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table A3.4
Descriptive Statistics: Teaching Practice

	Teaching Practice											
	Copying Notes			Having a Quiz/ Test			Giving Homework			Working in Groups		
	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy
Never	0.095	0.058	0.059	0.164	0.176	0.203	0.087	0.039	0.082	0.516	0.423	0.641
	(0.293)	(0.233)	(0.236)	(0.370)	(0.380)	(0.402)	(0.282)	(0.194)	(0.274)	(0.499)	(0.494)	(0.478)
Once in a while	0.427	0.347	0.302	0.576	0.582	0.557	0.448	0.489	0.422	0.312	0.371	0.254
	(0.494)	(0.476)	(0.459)	(0.493)	(0.493)	(0.496)	(0.496)	(0.500)	(0.493)	(0.462)	(0.483)	(0.434)
Pretty often/ always	0.477	0.596	0.639	0.260	0.242	0.240	0.464	0.472	0.496	0.173	0.206	0.105
	(0.499)	(0.491)	(0.480)	(0.438)	(0.428)	(0.426)	(0.498)	(0.499)	(0.499)	(0.378)	(0.405)	(0.306)
Observations	2214	1226	1335	2214	1226	1335	2214	1226	1335	2214	1226	1335

Notes: Data source: TIMSS 1999. This table reports descriptive statistics on teaching practices. Individual observations are weighted by sampling probabilities. Standard deviations in parentheses.

Table A3.5
Descriptive Statistics: Attitude toward Math

	Attitude											
	Like Math				Positive Attitude toward Math				Math Important for myself			
	Female		Male		Female		Male		Female		Male	
	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex
Low	0.479	0.450	0.426	0.448	0.309	0.271	0.244	0.253	0.091	0.092	0.105	0.106
	(0.500)	(0.498)	(0.495)	(0.498)	(0.462)	(0.445)	(0.430)	(0.435)	(0.288)	(0.288)	(0.307)	(0.308)
Medium	0.417	0.436	0.419	0.413	0.612	0.645	0.646	0.652	0.539	0.495	0.496	0.538
	(0.493)	(0.496)	(0.494)	(0.493)	(0.488)	(0.479)	(0.479)	(0.476)	(0.499)	(0.500)	(0.500)	(0.499)
High	0.105	0.114	0.155	0.138	0.079	0.084	0.110	0.095	0.369	0.413	0.399	0.356
	(0.306)	(0.318)	(0.362)	(0.345)	(0.270)	(0.277)	(0.313)	(0.293)	(0.483)	(0.493)	(0.490)	(0.479)
Observations	1120	1222	1087	1327	1114	1225	1089	1324	1117	1224	1089	1327

	Confidence in Math					Educational Aspiration			
	Female		Male			Female		Male	
	Coed	Single-Sex	Coed	Single-Sex		Coed	Single-Sex	Coed	Single-Sex
Low	0.060	0.051	0.040	0.040	Secondary Education	0.042	0.031	0.046	0.046
	(0.237)	(0.221)	(0.196)	(0.195)		(0.201)	(0.173)	(0.210)	(0.209)
Medium	0.848	0.873	0.833	0.846	Vocational Education	0.070	0.059	0.077	0.095
	(0.359)	(0.333)	(0.373)	(0.362)		(0.255)	(0.236)	(0.267)	(0.293)
High	0.092	0.076	0.127	0.115	Tertiary Education	0.794	0.804	0.770	0.745
	(0.290)	(0.265)	(0.333)	(0.319)		(0.405)	(0.397)	(0.421)	(0.436)
Observations	1117	1225	1089	1323		1116	1226	1085	1331

Notes: Data source: TIMSS 1999. This table reports descriptive statistics on student attitude toward math. Individual observations are weighted by sampling probabilities. Standard deviations in parentheses.

Table A3.6
Descriptive Statistics: Disciplinary Climate

	Behave Not Orderly			Behave Not as Told			Students Disruptive			Students Uninterested		
	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy
Not a Problem	0.129	0.089	0.133	0.080	0.060	0.087	0.204	0.386	0.241	0.282	0.435	0.372
	(0.333)	(0.284)	(0.336)	(0.269)	(0.236)	(0.280)	(0.403)	(0.487)	(0.428)	(0.450)	(0.496)	(0.484)
Minor Problem	0.590	0.511	0.594	0.517	0.430	0.494	0.530	0.498	0.575	0.429	0.433	0.476
	(0.489)	(0.499)	(0.488)	(0.497)	(0.494)	(0.497)	(0.499)	(0.500)	(0.495)	(0.495)	(0.496)	(0.500)
Serious Problem	0.281	0.400	0.273	0.403	0.510	0.419	0.266	0.116	0.184	0.289	0.132	0.152
	(0.447)	(0.489)	(0.444)	(0.488)	(0.499)	(0.491)	(0.442)	(0.321)	(0.388)	(0.453)	(0.339)	(0.359)
Observations	2214	1226	1335	2214	1226	1335	2167	1180	1335	2167	1180	1335
	Wide range of Backgrounds			**Wide range of Acad. Abilities**			**Intimidation of Students**			**Injury of Students**		
	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy	Coed	All-girl	All-boy
Not a Problem	0.290	0.500	0.481	0.182	0.487	0.580	0.348	0.352	0.363	0.429	0.693	0.363
	(0.454)	(0.500)	(0.500)	(0.386)	(0.500)	(0.494)	(0.477)	(0.478)	(0.481)	(0.495)	(0.461)	(0.481)
Minor Problem	0.566	0.416	0.440	0.616	0.407	0.359	0.499	0.548	0.543	0.493	0.233	0.513
	(0.496)	(0.493)	(0.497)	(0.486)	(0.491)	(0.480)	(0.500)	(0.498)	(0.498)	(0.500)	(0.423)	(0.500)
Serious Problem	0.144	0.084	0.079	0.202	0.106	0.060	0.153	0.100	0.094	0.078	0.074	0.124
	(0.351)	(0.278)	(0.269)	(0.402)	(0.308)	(0.238)	(0.360)	(0.300)	(0.292)	(0.268)	(0.262)	(0.330)
Observations	2167	1180	1335	2167	1226	1335	2135	1190	1335	2135	1190	1335

Notes: Data source: TIMSS 1999. This table reports descriptive statistics on the disciplinary climate. Individual observations are weighted by sampling probabilities. Standard deviations in parentheses.

Table A3.7
Descriptive Statistics: High School Students

	General High Schools Students				Vocational High Schools Students			
	Female		Male		Female		Male	
	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex	Coed	Single-Sex
Age	15.72	15.77	15.74	15.76	15.74	15.73	15.75	15.73
	(0.30)	(0.29)	(0.29)	(0.29)	(0.29)	(0.29)	(0.30)	(0.28)
Parents' Education: None	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.00
	(0.05)	(0.05)	(0.00)	(0.04)	(0.10)	(0.10)	(0.17)	(0.06)
Parents' Education: Primary	0.01	0.05	0.02	0.04	0.11	0.12	0.17	0.10
	(0.11)	(0.22)	(0.15)	(0.20)	(0.32)	(0.33)	(0.37)	(0.30)
Parents' Education: Secondary	0.61	0.62	0.62	0.60	0.70	0.71	0.63	0.69
	(0.49)	(0.49)	(0.49)	(0.49)	(0.46)	(0.45)	(0.48)	(0.46)
Parents' Education: University	0.38	0.32	0.36	0.35	0.18	0.15	0.18	0.20
	(0.48)	(0.46)	(0.48)	(0.47)	(0.39)	(0.36)	(0.38)	(0.40)
Books at Home: 0-10	0.03	0.04	0.02	0.04	0.19	0.16	0.18	0.17
	(0.16)	(0.19)	(0.13)	(0.20)	(0.39)	(0.36)	(0.38)	(0.38)
Books at Home: 11-50	0.11	0.14	0.12	0.14	0.24	0.28	0.30	0.21
	(0.31)	(0.34)	(0.32)	(0.34)	(0.43)	(0.45)	(0.46)	(0.41)
Books at Home: 51-100	0.21	0.23	0.22	0.21	0.26	0.24	0.23	0.27
	(0.41)	(0.42)	(0.41)	(0.41)	(0.44)	(0.43)	(0.42)	(0.44)
Books at Home: 101-250	0.33	0.32	0.31	0.31	0.17	0.20	0.17	0.25
	(0.47)	(0.47)	(0.46)	(0.46)	(0.38)	(0.40)	(0.38)	(0.43)
Books at Home: > 250	0.33	0.28	0.34	0.30	0.14	0.12	0.11	0.10
	(0.47)	(0.45)	(0.47)	(0.46)	(0.34)	(0.33)	(0.31)	(0.30)
Live with Parents	0.96	0.95	0.96	0.96	0.90	0.88	0.86	0.84
	(0.20)	(0.21)	(0.19)	(0.19)	(0.29)	(0.31)	(0.33)	(0.36)
Observations	413	643	409	1292	220.00	625.00	444.00	319.00

Notes: Data source: PISA 2000. Individual observations weighted by sampling probabilities. Standard deviations in parentheses.

Table A3.8
The Effect of Single-Sex Schooling at General High Schools

	Female								
	Math			Science			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Single-Sex Schooling	0.032	0.082	0.135	0.042	0.051	-0.072	0.046	0.070	0.002
	(0.099)	(0.088)	(0.117)	(0.123)	(0.116)	(0.125)	(0.086)	(0.081)	(0.092)
Student Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School Controls	No	No	Yes	No	No	Yes	No	No	Yes
Imp. Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	586	586	586	584	584	584	1056	1056	1056
Cluster	40	40	40	40	40	40	40	40	40
R^2	0.000	0.096	0.126	0.001	0.038	0.075	0.001	0.049	0.082
	Male								
	Math			Science			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Single-Sex Schooling	0.064	0.085	0.114	-0.014	0.002	0.061	0.040	0.055	0.058
	(0.089)	(0.082)	(0.092)	(0.094)	(0.084)	(0.093)	(0.086)	(0.080)	(0.088)
Student Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School Controls	No	No	Yes	No	No	Yes	No	No	Yes
Imp. Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	961	961	961	961	961	961	1726	1726	1726
Cluster	57	57	57	57	57	57	57	57	57
R^2	0.001	0.050	0.079	0.000	0.059	0.092	0.000	0.045	0.077

Notes: Data source: PISA 2000. Individual student observations are weighted by sampling probabilities. Standard errors are clustered at the school level and reported in parentheses. Student controls include age, parents' education, books at home and living with mother and father. School controls include total enrollment, school location, student-teacher- and computer-student ratios, hiring and course autonomy, shortage of math teacher, private institution and share of government funding. All regressions control for imputation. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Chapter 4

The Effect of Course Policies on Student Performance: Evidence from a Difference-in-Differences Approach*

4.1 Introduction

Improving the academic performance of students is a major concern of researchers and policy makers alike. Thus, substantial research has gone into understanding the determinants of educational achievement, mainly focusing on the role and importance of students' family backgrounds and school resources (see, e.g., Hanushek, 2003). However, students' outcomes depend not only on innate ability, family background and school resources, but to a large extent on individual characteristics such as motivation and effort. However, there is little research on student effort and the factors that influence it. This is surprising because it is common knowledge that performance in general depends on effort and there is no reason to believe that academic performance is an exception. Unlike other determinants of academic performance, such as innate ability, effort is both variable and susceptible

* This chapter was coauthored by Philipp Beltz, University of Munich and Andreas Ostermaier, Technische Universität München.

to incentives, which makes it a primary starting point for increasing the performance of students.

The existing literature on incentives and academic performance mostly focuses on the effects of monetary rewards, e.g., cash incentives, which are tied to academic progress. While there exists a common fear that monetary incentives reduce intrinsic motivation (e.g., Gneezy and Rustichini, 2000), there is evidence that rewards, non-monetary but also monetary, improve student achievement (see, e.g., Levitt et al., 2012). Among secondary school students, monetary rewards were found to increase performance, as measured by test scores and other outcomes such as completion rates (Angrist et al., 2002, 2006; Kremer et al., 2009; Angrist and Lavy, 2009). Among university students, there is evidence that merit-based scholarship programs and rewards raise enrollment rates (see, e.g., Cornwell et al., 2006) and that continuation fees prompt students to graduate within the scheduled time (see Garibaldi et al., 2012). However, financial rewards do not necessarily affect students' achievements (see Leuven et al., 2010). One explanation for this might be that students lack the knowledge to convert effort into measurable achievement (Fryer, 2011). Furthermore, mixed results on the effect of monetary rewards suggest that the design of incentives is crucial (Levitt et al., 2012).

This study investigates whether university program rules, such as credit points or the number of allowed resits, serve as incentives for students. In particular, we consider a business school at a German university that offers two similar study programs that both became subject to reforms. While the policies for the business administration program (*Betriebswirtschaftslehre*) were changed as early as 2005, the reform of the business education program (*Wirtschaftspädagogik*) was delayed until 2010. We analyze the effects of the modified program policies in a difference-in-differences approach and digitized students' performance data in a typical business course that is compulsory for both groups of students. The fact that both groups of students attend the same course, are taught by the same instructors, use the same textbooks and teaching materials, and that they have a nearly identical curriculum when they take the exam corroborates the common trend assumption that we need to make.

We find that the first reform, which effectively doubled the time until students receive their first certificate and which reduced the impact of each exam on the Grade Point Average, has a negative impact on student achievement. Furthermore, we show that a higher number of allowed resits increases the portion of students that submits blank papers. Since students must not resit exams once they have passed, those students failed deliberately to have the opportunity to resit the exam and improve. We also show that students respond differently to university policies depending on individual ability. Our results are robust to different specifications and robustness tests. In particular, we restrict our sample to groups of students that are less prone to selection and perform matching techniques to make students most comparable along the vector of observable characteristics. As an attempt to test the common trend assumption, we show that, conditional on students' age and semesters, both groups of students follow the same trend after being reformed in the same way.

To the best of our knowledge, there is no empirical evidence that program and course policies function as incentives. Moreover, and in contrast to monetary rewards, program policies must be adopted by universities and are generally inexpensive. Thus, they might be a promising tool of incentivising student effort.

The remainder of this chapter is organized as follows: Section 4.2 describes the institutional setting along with the reforms that we consider in this study. Section 4.3 depicts student responses to incentives in a theoretical setting. Section 4.4 describes the data. Section 4.5 presents the identification strategy and the empirical model. Sections 4.6 reports the baseline results for both reforms along with the subgroup analyses. Section 4.7 presents the robustness tests. Section 4.8 concludes with a discussion of our findings and their implications.

4.2 Institutional Setting

In 1999, the education ministers of 29 European countries agreed in Bologna to create the European higher Education Area, which required them to harmonize their national university systems. The so-called Bologna Process obliged German universities, which up to that point did not distinguish between Bachelor and Master degrees, to introduce sweeping reforms across their degree programs. These reforms affected grading, credit points, number of resits, and similar program and course policies.

This study focuses on the business school of a major public university in Germany, which offers undergraduate programs in business administration and business education.

Both programs give a broad knowledge of the functional areas of a company, although the business education program offers graduates the option of teaching at vocational schools in addition to applying for positions in the business sector. Until the Bologna reform, both programs were divided into two periods of study and students received the *Vordiplom* certificate at the end of the first period, and the *Diplom* certificate at the end of the second. At that time, graduates of both programs were rewarded what was known as the *Diplom* degree.

Official program policies specified which courses and exams students had to pass in order to earn their certificates and thus their degree. Both the *Vordiplom* and *Diplom* certificates reported an overall grade (hereafter Grade Point Average (GPA)), which averaged the grades students had obtained in the courses required for that certificate. While there was no *Vordiplom* degree, the *Vordiplom* certificate was a prerequisite to qualify for the second period of study. At the same time, the *Vordiplom* drew a line under the first period of study and grades earned until then did not count toward the *Diplom* certificate.

In order to adapt to the Bologna system, the business administration program was reformed in two steps, which involved a major reform in 2005 and a minor reform in 2008. In 2005, the four-year *Diplom* program was replaced with a three-year Bachelor program.[1] Since then, the Bachelor certificate is the first and only certificate students receive upon completing their program.[2] While the first three semesters of the Bachelor program are identical to the first three semesters of the *Diplom* program, the completion of the first half of the Bachelor program is no longer marked by anything equivalent to the *Vordiplom*. Also, in contrast to the earlier system, now all grades count toward the final Bachelor GPA.

In 2008, the new Bachelor program was revised and, as a result, the number of times students were allowed to resit an exam was increased for most courses. Before the revision, students (enrolled in either the *Diplom* program or the Bachelor program) were allowed a maximum of three attempts, whereas students enrolled in the revised Bachelor program

[1] In addition, a Master program and degree were introduced. Most of the courses to be taken in the last year of the *Diplom* program became part of the new Master program.

[2] Students can still retrieve transcripts of records at any time. However, this was also possible before the Bachelor program was introduced.

may now resit exams as often as they want as long as they graduate on time.[3] Since students are generally not allowed to repeat an exam once they have passed, this policy was not intended to give students the opportunity to improve their grades by resitting exams. Nevertheless, the higher number of allowed resits offers students more chances to eventually pass an exam. However, if students fail at their final attempt of an exam, they are not allowed to continue on this or a related program at *any* university in Germany. The main reason why the number of resits was limited until 2008 was to screen out students who were not suitable for the program of their choice. This forced students to realize early whether they had chosen the right program.

In 2010 the business education program was reformed in one single step, which comprised both reforms of the business administration program. Thus, the 2010 reform of the business education program restored the situation before 2005, when the organization of both programs was identical.

4.3 Theoretical Predictions

To get an idea how modifications of program and course policies may have an impact on student performance, a framework of student learning is presented in the following.

4.3.1 A Framework of Student Learning

Academic performance can be described as a function of family, peer, and school inputs as well as student characteristics, such as individual ability (Hanushek, 1986, 2002). The most extensive literature focuses on school resources, such as reducing class sizes (see, e.g., Hoxby, 2000). However, this input-based approach does not necessarily translate into student achievement gains (e.g., Hanushek, 1996, 2003). As a result, interest shifted towards incentives for the people involved in the process of education. Common examples include external exit exams and accountability policies, which create incentives for principals, teachers and individual students (e.g., Bishop, 1997; Bishop and Woessmann, 2004; Hoxby, 1994; Woessmann, 2003).

[3] The regular (maximum) study time is three (four) years. Given that students failed their previous attempts (otherwise they are not allowed to resit), they can resit the exam every semester, which adds up to eight attempts at most.

The main purpose of incentives for individual students is to increase their effort and, as a result, performance. This approach relies on the intuitive assumption that performance depends critically on effort. In support of this assumption, the OECD's Programme for International Student Assessment revealed that truancy and inattention, which can be taken to reflect a lack of effort, correlate with poor reading and mathematical skills (Bishop, 2006); conversely, attendance of tutorials was found to enhance performance (Durden and Ellis, 1995). Likewise, in some studies it was possible to attribute the effects of monetary rewards to increased effort (Angrist et al., 2002, 2009). Accordingly, effort is included as an input in certain education production functions (e.g., Akerlof and Kranton, 2002; Bishop and Woessmann, 2004; Bishop, 2006).[4]

Unlike other determinants of academic performance, including innate ability and family background, effort is controllable by the student. Following Akerlof and Kranton (2002), students choose the level of effort that maximizes their expected net benefit, that is, the difference between the benefits and costs of studying. Overall, students can be intrinsically or extrinsically motivated to learn. In the former case, they find learning itself rewarding, whereas in the latter, they consider it as a means of obtaining other rewards, such as recognition or the prospect of higher earnings. Consequently, the benefits consist of intrinsic and extrinsic rewards, while the costs may be monetary or non-monetary (e.g., tuition fees, but also time, strain, stress, etc.).

4.3.2 Student Effort Choices

By increasing the benefits of studying, monetary rewards are thought to enhance student effort and thus performance. Similarly, program and course policies may function as incentives if they influence the perceived costs and benefits of studying.

Timing of Reward

As rational actors, students will account for time when comparing the costs and benefits of studying for an exam. Although studying may be experienced as intrinsically beneficial, benefits such as a better job or higher lifetime earnings are obtained in the future, while students have to put in effort and pay for the costs now. Hence, they will discount the

[4] However, there exists evidence that students lack the ability to translate effort into measurable achievement (Fryer, 2011).

expected benefits to the "net present value" in order to compare them with the costs (Frederick et al., 2002).

In this vein, it has been argued that monetary rewards for students might correct for high discount rates (Angrist and Lavy, 2009). Additionally, Levitt et al. (2012) find that only immediate incentives work which suggests that the timing of the reward matters.

Impact on Reward

When choosing their effort levels, students also account for the impact their effort might have on their certificates. Since some courses require more time and effort than others, depending on the complexity of the course contents, universities account for these differences by calculating the number of credit hours or points.[5] While the number of credit points reflect objective differences between courses, they are at the same time the weight of the grade in the GPA. Given the limitation of resources such as time, a higher number of credit points may animate students to focus on the respective course, because it offers more "leverage" to the performance that should result from their effort. Consequently, students will put more effort in courses with more credit points.

More generally, students will not only consider the number of credit points of one course as compared to others when they choose their levels of effort, but also the weight of the course as such. The larger the number of courses factored into the GPA, the smaller the weight of each course. Consequently, even a course that carries more credit points than others may offer little perceived leverage.

Cost of Failing

Students may also consider the consequences of failing while studying for an exam. Imagine that students either have only one attempt to pass an exam or students can resit their exams as often as they want as long as they finish their studies within the required time. If students do not have the chance to resit an exam, the cost of failing is prohibitive or, put differently, the benefit of succeeding is immense. Those who fail have to leave their programs and in some cases may even not be allowed to continue on the same program at

[5] In the European Credit Transfer System (ECTS), "credit points" corresponds to what is more commonly known as "credit hours" in the U.S. As with credit hours, the number of credit points of a course is supposed to reflect the time students have to spend on it.

a different university. In contrast, if students can resit exams as often as they want, the cost of studying remains the same, but the benefit of succeeding decreases for all attempts except the final one.

On the one hand, students may generally put in less effort and prepare worse for an exam if they know that they have the opportunity to resit an exam several times. On the other hand, if students are not allowed to resit an exam unless they have failed, they may decide to fail deliberately rather than earning a bad grade. Those students will not answer at all or they cancel their answers before submitting their exams.

4.3.3 Expectations

The framework outlined above suggests that students will adjust their effort levels if the costs and benefits of studying change. Thus, unless students are exclusively driven by intrinsic motivation, modifications of the program and course policies that shift students' cost–benefit ratios should have an impact on their effort choices. If effort translates into performance, modifications of the program and course policies should have an impact on student achievement.

2005 Reform

The reform of 2005 replaced the *Diplom* program with the Bachelor program for business administration students which effectively increased the time until students receive their first certificates and decreased the leverage of each exam on their GPAs. In particular, students enrolled in the *Diplom* degree earned the *Vordiplom* certificate after 1.5 years, while students enrolled in the Bachelor degree receive their first and only certificates after at least 3 years. Since the Bachelor certificate reports all grades obtained during the program, the courses that count for the Bachelor GPA are more than twice as many as those that counted for the *Vordiplom* GPA. For example, a typical exam taken in the first three semesters of the business administration program accounted for about seven percent of the GPA in the *Vordiplom* certificate and for approximate three percent of the GPA in the Bachelor certificate.

Certificates are the most important rewards of studying, because they serve as a "signal" when students apply at the job market, for scholarships and for programs at other universities (Hanushek, 2002; Spence, 1973). While the *Vordiplom* certificate was not a degree, it was nevertheless used to apply for internships, study abroad semesters and scholarships. In addition, employers usually asked for the *Vordiplom* certificate in the application procedure. Furthermore, and despite the fact that the Bachelor is the first academic degree relevant to the labor market, most bachelor students actually plan to enroll in a master's degree after graduating (Grützmacher et al., 2011).[6]

This suggests that the *Vordiplom* certificate and the Bachelor certificate are quite comparable regarding their signaling effects, but students receive their Bachelor certificates only after three years. As students discount future rewards, the increased time until students can use their certificates as signals, might have negative implications for motivation and effort choices (time effect). In addition, the introduction of the Bachelor degree reduced the impact of every grade obtained during the first period of study on the GPA (leverage effect). Consequently, we expect that the 2005 reform had a negative impact on student performance.

2008 Reform

The 2008 reform increased the number of allowed resits. Because students are generally not allowed to repeat an exam once they have passed, they need to fail if they want to resit an exam in order to improve. Given that there is to some extent randomness in every exam, due to the topics covered in the exam (students may have studied selectively due to, e.g., time constraints) or the individual situation (such as unforeseen sickness or lack of concentration), we expect that the higher number of allowed resits increased the share of students that failed deliberately by submitting blank exams. Furthermore, students may study less serious given the opportunity to resit exams several times. Thus, we expect that student performance even decreased for students that did not submit blank.

4.4 Data

To analyze the effects of the modified course policies, we collected data on the business administration and business education students between 2006 and 2012. In particular, we

[6] In 2009/2010, 55 percent of bachelor students aimed to do a master's degree and 27 percent considered it (Grützmacher et al., 2011).

observe the performance of students in a typical business course that is recommended to be taken in students' third semesters and is compulsory for both business administration and business education students, no matter whether the former were in the *Diplom*, the Bachelor, or the revised Bachelor programs. The course covers topics on production and controlling and is assessed in a one-hour written exam that requires both analytical and quantitative skills. Since the course is offered every semester, it is lectured by several chairs in a rotating manner. By focusing on the exams of 2006, 2008, 2010, and 2012 we make sure that we observe students who were taught by the same lecturer who covered the same topics and problem sets. Our data cover four generations of students, which come to 1,630 observations.

We compiled the data from two sources within the university. The office of the university registrar collects personal data from students when they apply for admission, such as age, gender, country of birth, or qualifications obtained before enrollment. The office of the registrar of the business school keeps academic records and files exams. However, academic records only report grades, whereas we wanted to rely on the more finely partitioned test scores. We therefore retrieved the exams from the file room and digitized the data on performance for each observation. In order to combine information on performance with information on personal characteristics, we finally matched the data from these sources by using a student's unique registration number.

We consider only students enrolled in the business administration or business education programs. We therefore excluded exchange students, because they are not subject to the program and course policies examined in this study, and nine observations for which we do not have information on their programs. In addition, two students enrolled in the *Diplom* program took the exam only in 2008 (together with those in the Bachelor program) rather than 2006, as well as 16 in the Bachelor program and one in the *Diplom* program only in 2010 rather than 2008 and 2006 (together with those in the revised Bachelor program), respectively. These observations were also discarded.

In addition to the test scores, which range from 0 to 120 points, and the programs students were enrolled in, our database contains information on demographic characteristics including gender, age, and country of birth. It also includes information on the number of semesters a student had officially been enrolled in his or her program when he or she took the exam ("semester"). We also managed to access students' grades of their secondary education certificates ("high school GPA"), which we rescaled such that higher numbers indicate higher performance. We define exams scoring ten points or less as "blank" to indicate students that failed deliberately. Among the students that actually scored zero, some students did not answer at all and some cancelled all their answers before submitting. Nevertheless, we set the threshold at ten points, because students that scored only slightly above zero, were presumably sure enough that they will not pass or they just forgot to cancel some of their answers. Overall, we miss information on age for one percent of the students and information on the high school GPA for ten percent of the students.

Table 4.1 shows that in each year between 60 and 80 percent of the students were enrolled in the business administration program, the rest in the business education program. The mean test score varies considerably between the exams, ranging from slightly above 60 in 2010 to around 80 in 2012. The rate of failure was about 25 percent in 2006, but decreased to about 15 percent in the following years. The rate of blank submission rose strikingly from zero in 2006 to six percent in 2010 in business administration and remained above zero in 2012 for both programs. The share of female students is about 50 percent among business administration students and between 60 and 70 percent among business education students. The proportion of students born abroad ranged from four to eleven percent in business administration, from zero to three percent in business education. The high school GPAs were about 3.0 ("B"), with major increases from 2008 to 2010 among business administration students, and from 2010 to 2012 among business education students. Most students took the exam, as it was recommended, in their third semesters at university, which is at about the age of twenty-three in business administration, and twenty-four in business education. The business education students were on average older because many of them received vocational training before enrolling at university.

Unfortunately, students who enrolled in business education in 2012 received the recommendation to take the exam considered in this study in their first semesters. As a result, the universe of business education students in 2012 is generally younger and takes the exam in lower semesters (see Table 4.1). This suggests that the composition of the two groups of students is not fully comparable in 2012. Nevertheless, we use the data of 2012 for our robustness test while accounting for students' age and semesters.

Table 4.1
Descriptive Statistics

	2006			2008			2010			2012		
	BA (1)	BE (2)	Diff. (3)	BA (4)	BE (5)	Diff. (6)	BA (7)	BE (8)	Diff. (9)	BA (10)	BE (11)	Diff. (12)
Test score	72.41	62.59	8.76	79.77	79.70	-1.33	58.53	60.36	-2.98	79.87	78.86	1.74
	(20.16)	(14.77)	(2.84)	(18.39)	(14.42)	(1.99)	(21.99)	(17.72)	(2.01)	(23.07)	(22.18)	(2.58)
Blank	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.00	0.05	0.03	0.03	-0.01
	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)	(0.23)	(0.00)	(0.01)	(0.17)	(0.17)	(0.02)
Failure	0.23	0.33	-0.08	0.13	0.08	0.06	0.19	0.17	0.05	0.14	0.14	-0.01
	(0.42)	(0.48)	(0.08)	(0.33)	(0.28)	(0.04)	(0.40)	(0.38)	(0.04)	(0.35)	(0.35)	(0.04)
Male	0.53	0.33	0.17	0.54	0.36	0.15	0.47	0.30	0.15	0.53	0.32	0.21
	(0.50)	(0.48)	(0.09)	(0.50)	(0.48)	(0.06)	(0.50)	(0.46)	(0.05)	(0.50)	(0.47)	(0.05)
Foreign	0.04	0.00	0.10	0.09	0.03	0.12	0.11	0.02	0.15	0.08	0.02	0.06
	(0.19)	(0.00)	(0.04)	(0.28)	(0.16)	(0.03)	(0.32)	(0.16)	(0.03)	(0.27)	(0.14)	(0.02)
High School GPA	2.85	2.67	0.17	2.97	2.78	0.18	3.11	2.71	0.40	3.02	2.92	0.10
	(0.55)	(0.67)	(0.12)	(0.45)	(0.53)	(0.07)	(0.50)	(0.51)	(0.05)	(0.44)	(0.45)	(0.05)
Age	22.74	25.51	-2.62	22.84	24.36	-1.43	22.69	24.39	-1.66	22.96	22.64	0.29
	(1.43)	(3.72)	(0.60)	(2.82)	(2.38)	(0.31)	(2.09)	(2.60)	(0.27)	(2.45)	(2.56)	(0.28)
Semester	2.95	3.26	-0.28	3.01	3.14	-0.12	2.98	3.11	-0.13	3.22	2.20	1.01
	(0.63)	(0.99)	(0.16)	(0.18)	(0.38)	(0.05)	(0.24)	(0.64)	(0.06)	(0.88)	(0.98)	(0.11)
Observations	183	54	237	326	80	406	313	123	436	452	99	551
	[129]	[39]	[168]	[288]	[73]	[361]	[285]	[122]	[407]	[428]	[98]	[526]

Notes: This table reports the means on observable characteristics of business administration students (BA) and business education students (BE) along with the difference between the two groups in 2006, 2008, 2010, and 2012. Standard deviations in parentheses. The number of all observations including those with missing values appears in parentheses, the number of observations for which complete data is available appears in brackets below.

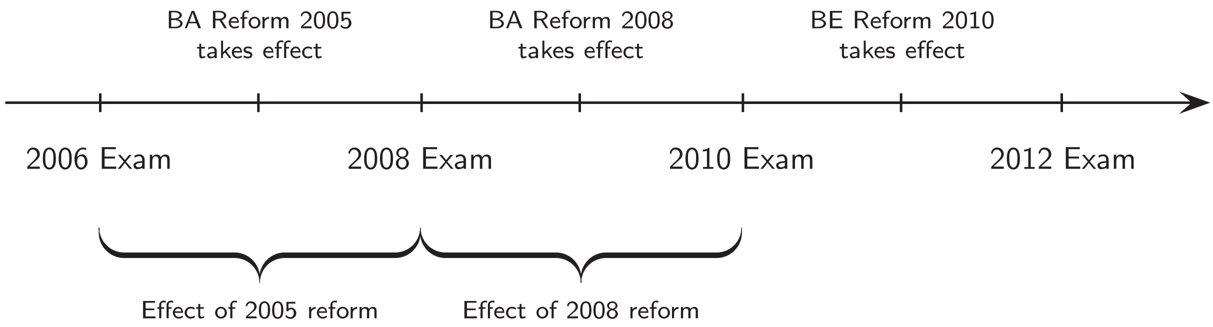
4.5 Identification Strategy and Empirical Model

We examine the effects of the two reforms in a quasi-experimental setting, which resulted from the Bologna reform of higher education in Europe. Since not all universities adopted the Bologna system at the same time, and even within universities some programs were revised later than others, it happened that at some universities the traditional and the reformed systems coexisted temporarily. In particular, we consider a business school that offers two similar programs of study, which both became subject to reforms. While the program policies changed twice for business administration students since 2005, they remained the same for business education students until 2010.

Figure 4.1 presents a timeline of the data described above along with these reforms. The business administration students who took the exam in 2006 were enrolled in the *Diplom* program, those who took it in 2008, in the Bachelor program (which was introduced by the 2005 reform), and those who took it in 2010, in the revised Bachelor program (which was introduced by the 2008 reform). Conversely, during the entire period between 2006 and 2010, the business education students were enrolled in the *Diplom* program. In 2012, however, both groups of students were in Bachelor programs, and as noted above, the organization of the newly introduced Bachelor program in business education corresponded to the revised Bachelor program in business administration.

We consider business administration students, who were subject to the reforms, as treatment group and business education students as control group. We then compare the development of business administration students with the development of business education students in a difference-in-differences framework. In particular, we use the assessment data of 2006 (2008), which refers to a point in time before the 2005 (2008) reform *came into effect* for business administration students and the assessment data of 2008 (2010), which refers to a point in time after the reform *came into effect* (see also Figure 4.1). This approach allows us to separate the effects of both reforms, since there is only one change in program policies for business administration students between the years considered. In addition, we check the consistency of our results and report the effect of the revised Bachelor program, that comprises the features of the 2005 and the 2008 reforms, by using the data of 2006 and 2010 only. Furthermore, the results for a pooled regression using the data of 2006, 2008, and 2010 are presented.

Figure 4.1. Timeline of the Reforms.



Notes: The business administration (BA) students, who took the exam in 2006, were enrolled in the *Diplom* program, those who took it in 2008, in the Bachelor program, and those who took it in 2010, in the revised Bachelor program. Conversely, during the entire period between 2006 and 2010, the business education (BE) students were enrolled in the *Diplom* program. In 2012, however, both groups of students were in Bachelor programs.

We consider three measures of performance as outcome variables, namely (i) test scores, (ii) the rate of failure, and (iii) the rate of blank submission and estimate the following baseline model:

$$Y_i = \alpha + \beta_1 P_i + \beta_2 T_i + \delta P_i T_i + \gamma' X_i + \varepsilon_i. \tag{4.1}$$

Y_i is student i's performance (test score, rate of failure, and blank submission), while the dummy variables P_i and T_i indicate a student's field of study (1, if business administration) and the year of examination (1, if after the respective reform). $P_i T_i$ is the interaction of the dummy variable of the year of examination and the dummy variable of the program of business administration. X_i denotes a set of control variables, and ε_i an idiosyncratic error term. In this model, δ, captures the effect of the reform, β_1 the time-invariant effect specific to the business administration students, β_2, the time-effect for the control group, and the vector γ includes the coefficients of the control variables, namely gender, migration, high school GPA, age, and semester.

We report estimates with and without the set of control variables. Our baseline specification will be the simple comparison between control and treatment group, not controlling for any covariates. In order to correct for remaining differences between the treatment and control group, we add control variables to the regression. As a side effect, this reduces the residual variation and improves the precision of our estimates. First, we add controls for characteristics of the student, which are determined before students enroll at university, i.e. gender, migrant status, and high school GPA. Then, we additionally control for student's age and semester. As these variables could be endogenously related to the treatment, we account for this in our robustness section.

Interpreting δ as a causal treatment effect rests on two key identifying assumptions. First, we have to assume that test scores of both groups would have emerged with the same trend in absence of the reforms. The setting that we consider corroborates this assumption and suggests that the groups of students are indeed comparable: To begin with, we measure performance in a typical business course that is compulsory for both groups of students. In particular, business education and business administration students in a given year (either 2006, 2008, 2010, or 2012) attend exactly the same course (they might actually sit next to each other), which is taught by the same lecturer, and covers the same teaching material and textbooks. At the end of the semester, both groups of students write the identical exam under identical circumstances. For example, a barking dog outside would affect both groups of students in the same way. Moreover, the lecturer and course content of this particular course did not change over time, such that all students, independent whether they have written the exam in 2006, 2008, 2010, or 2012, were taught by the same lecturer who covers the same topics and problem sets. Furthermore, by the time that the students wrote the exam, both groups of students have a nearly identical academical curriculum which should be translated into the same workload and background knowledge. Additionally, they lived in the same city, shared similar social environments, and had similar career opportunities.

The second assumption we have to make is that the composition of students within both groups did not change as a result of the treatment. In particular, students might have preferred the well-known *Diplom* to the yet unfamiliar Bachelor program. Since the admission to the *Diplom* program was closed when the Bachelor program was introduced at the university considered in this study, new students had no choice but to enroll into the latter. Furthermore, students had difficulties to avoid the Bachelor programs by enrolling at a different university since the *Diplom* programs in business administration were replaced with Bachelor programs at many comparable universities in Germany at about the same time.[7] In addition, students could enroll in business education rather than business administration, which implies that they would have chosen a different program to avoid the Bachelor. However, while our research benefits from the similarity between both programs, prospective students did not generally know about it and thus would not easily switch to the other program.

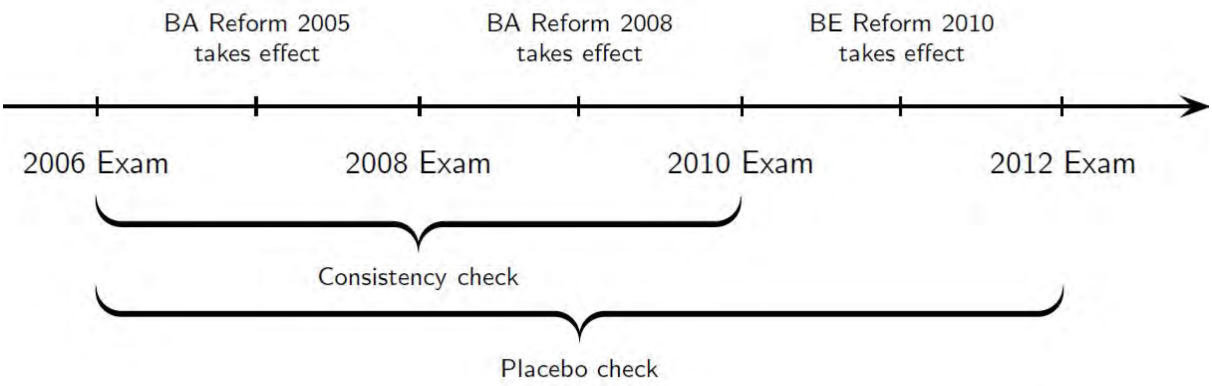
The descriptive statistics presented above provide an indication of the comparability of the two groups. As can be seen by Table 4.1, there are differences in observable student characteristics between the two groups in 2006, i. e. before the reforms came into effect. These differences are not necessarily a threat to identification as long as these differences remain constant over time. If there has been a selection into the Bachelor program, this should be reflected in the observable statistics of business administration students in 2008. However, except for age, the differences in 2006 and 2008 between business administration and business education students are nearly identical. Furthermore, most of the differences in 2010 are very similar to those in 2008 and 2006. However, the difference in high school GPAs almost doubled. Since the 2008 reform involved a very small change in the program rules (the increase in the number of allowed resits), it is rather unlikely that the increase in the high school GPAs of business administration students is as a result of this minor reform.

While we cannot fully explain this development, we know that applications increased faster than admissions over the period considered. As a result, the number of students as well as their high school GPAs should have increased in both programs. The biggest threat arising from differences in the development of the GPAs across groups is the self-selection of students based on their ability into the reformed programs. If student ability affects both treatment status and performance, the estimates of the reform effect will necessarily be biased. In order to alleviate this threat, we control for high school GPA, which is the most informative indicator of the overall ability to study of German students (Hell et al., 2007). Furthermore, we show that our results are robust to groups of students that are less prone to selection. In particular, we restrict our sample to students in their third semester and students that are in the 90^{th} percentile of the age distribution. Moreover, we

[7] Unfortunately, we could not find official documents that report exact dates for the introduction of the business adminstration Bachelor programs by German universities. However, by calling the offices of registrar, we found out that the business administration Bachelor program was introduced in the winter semester 2004/2005 at the Humboldt University in Berlin and in the winter semester 2005/2006 at the Goethe University in Frankfurt.

perform matching techniques to make students most comparable along the full vector of observable characteristics. As an attempt to test the common trend assumption, we show that, conditional on age and semester, students respond in the same way to the reforms by comparing the exams of 2006 and 2012 (see Figure 4.2).

Figure 4.2. Robustness Checks.



Notes: The business administration (BA) students, who took the exam in 2006, were enrolled in the *Diplom* program, those who took it in 2008, in the Bachelor program, and those who took it in 2010, in the revised Bachelor program. Conversely, during the entire period between 2006 and 2010, the business education (BE) students were enrolled in the *Diplom* program. In 2012, however, both groups of students were in Bachelor programs.

4.6 Results

4.6.1 Effects of the 2005 Reform

Table 4.2 reports the effects of the 2005 reform according to the difference-in-differences model described above. In particular, we use assessment data of 2006, which refers to a point in time before the 2005 reform came into effect for business administration students and assessment data of 2008, which refers to a point in time after the reform came into effect. We investigate the effect of the 2005 reform only on test scores and the rate of failure, since we do not expect any effect on blank submissions and, as can be seen from Table 4.1, the shares of blank submissions are zero for both years and programs.

The left-hand side of Table 4.2 shows that the 2005 reform had a significant, negative effect on test scores. In particular, column (1) shows a negative coefficient for the universe of students, which remains about the same size and significance if we restrict the analysis

Table 4.2
Effects of the 2005 Reform

	Test Score				Failure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform 2005	-8.947***	-9.751***	-9.578***	-8.436***	0.144*	0.144	0.140	0.105
	(3.118)	(3.557)	(3.132)	(3.191)	(0.083)	(0.093)	(0.087)	(0.086)
2008	16.642***	17.109***	15.745***	14.723***	-0.270***	-0.251***	-0.232***	-0.205***
	(2.548)	(2.885)	(2.542)	(2.619)	(0.074)	(0.082)	(0.077)	(0.077)
Bus. Admin.	7.526***	9.825***	7.491***	4.955*	-0.093	-0.101	-0.063	0.006
	(2.441)	(2.940)	(2.675)	(2.852)	(0.074)	(0.084)	(0.079)	(0.081)
Male			0.700	0.926			-0.038	-0.045
			(1.379)	(1.364)			(0.030)	(0.029)
Foreign			-16.141***	-15.598***			0.247***	0.233***
			(3.703)	(3.736)			(0.083)	(0.084)
High School GPA			16.186***	15.056***			-0.228***	-0.193***
			(1.583)	(1.546)			(0.032)	(0.031)
Age				-1.138***				0.022***
				(0.355)				(0.007)
Semester				1.587				0.040
				(1.741)				(0.042)
Constant	62.870***	62.590***	19.070***	45.890***	0.370***	0.333***	0.957***	0.175
	(1.968)	(2.343)	(4.963)	(11.861)	(0.066)	(0.076)	(0.122)	(0.262)
Observations	644	529	529	529	644	529	529	529
R^2	0.070	0.073	0.289	0.314	0.042	0.037	0.153	0.176

Notes: This table reports OLS estimates of the effect of the 2005 reform on test scores and the rate of failure. The variable "Reform 2005" is an interaction of the 2008 year dummy variable and the business administration program dummy variable. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

to students for which we have complete information on personal characteristics (column 2). We then additionally control for variables that have been determined before students start their university career, namely gender, migrant status and high school GPA. As can be seen from column (3), this does not affect the coefficient of interest. In column (4), we additionally control for age and semester. The size of the coefficient drops a tiny bit, which is due to the fact that we control for age, a variable for which the difference between the two groups increases over time. We additionally performed this analysis controlling for age and semester separately, which does not change the overall result.

The right-hand side of Table 4.2 reports the effect of the 2005 reform on the rate of failure. As the test scores decrease due to the reform, the coefficient on the reform variable is positive and slightly significant in column (5). Restricting the analysis to students with nonmissing data (column 6) and controlling for predetermined student characteristics (column 7) does not change the coefficient of interest, although the significance vanishes due to slightly increased standard errors. After controlling for age and semester, the coefficient drops from about 14 percent to 10 percent in column (8). This is due to the fact, that older students are more likely to fail and that business education students were older on average in 2006 than in 2008.

Overall, Table 4.2 shows that the 2005 reform had a significant, negative effect on test scores and a positive effect on the probability to fail of business administration students, which is however not statistically significant.

Heterogeneous Effects

To check whether the effects of the 2005 reform are heterogeneous with respect to ability, we divide our students into terciles by using their high school GPA and refer to them as poor, average and excellent students. Since the high school GPA is not continuous and we have disproportionately more mass at the tercile points, our groups are not of exact equal sizes.

Table 4.3 reports the effect of the 2005 reform on test scores and the rate of failure by high school GPA. The results suggest that the 2005 reform caused average and excellent students to score lower (columns 3 and 5). While this translates into a positive effect on the rate of failure for average students (column 4), which is large in magnitude and statistically significant at the 10 percent level, the effect on the rate of failure is close to

Table 4.3

Effects of the 2005 Reform by High School GPA

	High School GPA					
	Low (≤ 2.7)		Average (> 2.7 & ≤ 3.1)		High (> 3.1)	
	Test Score	Failure	Test Score	Failure	Test Score	Failure
	(1)	(2)	(3)	(4)	(5)	(6)
Reform 2005	-4.968	0.125	-15.456**	0.328*	-12.962**	0.008
	(5.196)	(0.154)	(7.135)	(0.183)	(5.673)	(0.103)
2008	10.732**	-0.221*	22.168***	-0.366**	16.808***	-0.064
	(4.217)	(0.131)	(5.766)	(0.166)	(4.323)	(0.093)
Bus. Admin.	0.490	0.061	12.155*	-0.228	10.273**	-0.006
	(4.449)	(0.134)	(7.023)	(0.187)	(5.074)	(0.110)
Male	2.727	-0.093	1.152	-0.016	-1.746	-0.010
	(2.696)	(0.066)	(2.290)	(0.051)	(2.167)	(0.031)
Foreign	-5.734	0.295	-23.406***	0.335**	-15.678***	0.192*
	(7.321)	(0.230)	(6.046)	(0.134)	(5.816)	(0.114)
High School GPA	6.462	-0.136	24.016**	-0.300	18.033***	-0.040
	(5.269)	(0.124)	(9.793)	(0.203)	(5.624)	(0.074)
Age	-1.433***	0.028***	-1.149**	0.020	-0.602	-0.003
	(0.500)	(0.008)	(0.523)	(0.015)	(0.634)	(0.011)
Semester	-2.728	0.131**	11.123**	-0.240**	5.578	-0.089
	(2.208)	(0.057)	(4.857)	(0.108)	(3.916)	(0.067)
Constant	89.160***	-0.395	-15.375	1.506**	9.953	0.548
	(23.822)	(0.502)	(36.726)	(0.752)	(26.967)	(0.438)
Observations	187	187	172	172	170	170
R^2	0.154	0.148	0.282	0.174	0.224	0.096

Notes: This table reports OLS estimates of the effect of the 2005 reform on test scores and the rate of failure by high school GPA. The variable "Reform 2005" is an interaction of the dummy variable for the year 2008 and the dummy variable of the business administration program. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

zero and not statistically significant among excellent students (column 6). Among poor students, we find the same pattern, the coefficients are, however, smaller and statistically not significant.

This finding is consistent with the idea that the impact of effort on performance depends on ability. Poor students may have difficulties to control their performances via their effort choices, because they have trouble to concentrate or motivate themselves, regardless whether they study under the old or the new program rules. In contrast, good students know that they have the ability to influence their performances. If they make more or less effort, their performances normally increases or decreases. This interpretation suggests that the decrease in test scores is due to lower effort choices of average and excellent students.

4.6.2 Effects of the 2008 Reform

Tables 4.4 and 4.5 report the effects of the 2008 reform according to the difference-in-differences model described above. In particular, we use assessment data of 2008, which refers to a point in time before the 2008 reform came into effect for business administration students, and assessment data of 2010, which refers to a point in time after the reform came into effect. As the reform increased the number of allowed resits, we first investigate if this increased the share of students that deliberately failed by submitting blank exam sheets. We also investigate the effect on test scores and failure rates while accounting for blank submissions in the regression.

Table 4.4 shows that the 2008 reform increased the rate of blank submissions by about four to five percent. The coefficient is robust to restricting the sample to observations with nonmissing values and controlling for student characteristics (columns 2 to 4). Students that submitted blank have by definition a test score below ten points, thus they necessarily failed. In order to investigate whether the reform had an effect on test scores and failure rates of those that did not submit blank, we control for students that submitted blank in the regression. Table 4.5 shows that the performance of those that did not submit a blank exam has not been affected by the opportunity to resit this particular exam more often.

Table 4.4
Effects of the 2008 Reform on Blank Submission

	Blank Submission			
	(1)	(2)	(3)	(4)
Reform 2008	0.045***	0.053***	0.055***	0.055***
	(0.013)	(0.014)	(0.014)	(0.014)
2010	-0.000***	0.000***	-0.001	-0.001
	(0.000)	(0.000)	(0.002)	(0.002)
Bus. Admin.	0.006	0.003	0.004	0.003
	(0.004)	(0.003)	(0.005)	(0.006)
Male			-0.000	0.000
			(0.011)	(0.011)
Foreign			0.040	0.041
			(0.032)	(0.032)
High School GPA			-0.014	-0.015
			(0.013)	(0.014)
Age				-0.001
				(0.001)
Semester				0.002
				(0.003)
Constant	0.000***	-0.000***	0.039	0.048
	(0.000)	(0.000)	(0.036)	(0.064)
Observations	842	768	768	768
R^2	0.025	0.032	0.039	0.039

Notes: This table reports OLS estimates of the effect of the 2008 reform on the rate of blank submissions. The variable "Reform 2008" is an interaction of the dummy variable for the year 2010 and the business administration program dummy variable. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 4.5
Effects of the 2008 Reform on Test Scores and the Rate of Failure

	Test Score				Rate of Failure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform 2008	1.212	1.288	-1.505	-1.343	-0.041	-0.067	-0.033	-0.037
	(2.697)	(2.766)	(2.397)	(2.395)	(0.056)	(0.055)	(0.053)	(0.052)
2010	-19.212***	-19.338***	-18.263***	-18.336***	0.071	0.090*	0.076*	0.079*
	(2.269)	(2.323)	(1.951)	(1.946)	(0.048)	(0.047)	(0.045)	(0.043)
Bus. Admin.	-1.053	0.284	-1.970	-3.172*	0.045	0.040	0.068*	0.102***
	(1.916)	(1.986)	(1.669)	(1.691)	(0.039)	(0.038)	(0.035)	(0.035)
Blank	-59.932***	-60.641***	-57.500***	-57.623***	0.828***	0.856***	0.809***	0.812***
	(1.544)	(1.436)	(2.720)	(2.586)	(0.020)	(0.020)	(0.041)	(0.038)
Male			1.435	1.725			-0.031	-0.037
			(1.174)	(1.170)			(0.024)	(0.024)
Foreign			-9.614***	-9.312***			0.177***	0.167***
			(2.706)	(2.713)			(0.056)	(0.056)
High School GPA			13.811***	12.250***			-0.175***	-0.136***
			(1.370)	(1.411)			(0.028)	(0.029)
Age				-0.914***				0.022***
				(0.341)				(0.006)
Semester				-0.241				0.056
				(2.172)				(0.049)
Constant	79.513***	79.699***	40.989***	68.253***	0.100***	0.082**	0.577***	-0.241
	(1.619)	(1.682)	(4.212)	(10.800)	(0.034)	(0.032)	(0.092)	(0.231)
Observations	842	768	768	768	842	768	768	768
R^2	0.361	0.382	0.481	0.491	0.108	0.129	0.196	0.220

Notes: This table reports OLS estimates of the effect of the 2008 reform on test scores and the rate of failure. The variable "Reform 2008" is an interaction of the dummy variable for the year 2010 and the business administration program dummy variable. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Heterogeneous Effects

Again, we divide our students into terciles by using their high school GPA to investigate heterogeneous effects and refer to them as poor, average and excellent students. Table 4.6 shows a positive and significant effect of the 2008 reform on the rate of blank submissions among average and excellent students, with the coefficient being larger for average than for excellent students (columns 4 and 7). Moreover, the 2008 reform had a negative effect on test scores of average students that did not submit a blank exam (column 5), while there is no effect on the test scores of excellent students (column 8). Among poor students, there are no significant effects on blank submissions, test scores and the rate of failure.

This finding is also intuitive. Given that poor students have difficulties in preparing exams, their "objective functions" probably differ from that of average and excellent students. They tend to answer the exam questions, hoping that they may have just passed or done sufficiently well, rather than submitting a blank paper deliberately. In contrast, average and excellent students have the potential and goal to receive good results. However, there is to some extent randomness in every exam. While excellent students may manage to prepare themselves for several exams, average students might learn selectively and the exam may happen to cover unprepared topics. But even if students have prepared themselves very well, it is possible that they do not feel well or have suddenly problems to concentrate. Thus, both groups of students will submit a blank paper if they expect to do significantly better at a resit. However, the probability to do so is larger among average students.

Long-term Effects

Tables 4.4 and 4.6 suggest that the average number of attempts to pass an exam increases if students are allowed to resit exams more often. This might be beneficial if resits improve students' levels of training and increase their likelihood of success. Thus, it would at least help those students who finally pass and would have dropped out otherwise. However, resits may as well decrease motivation, which leads students to perform even worse as they retake the exam more often. Moreover, the number of exams to be taken at a time increases as students procrastinate, which makes success even more unlikely. It is therefore

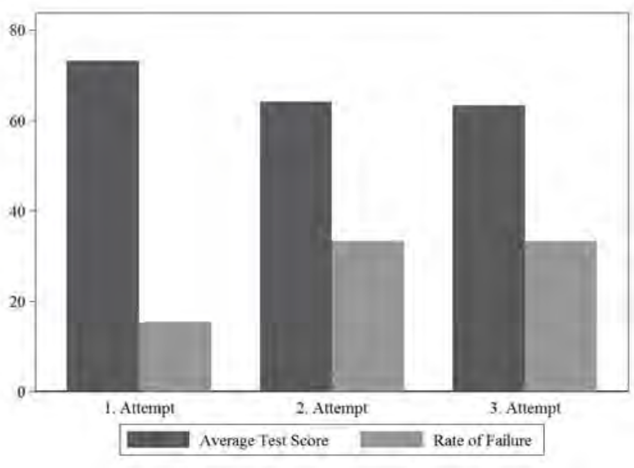
Table 4.6
Effects of the 2008 Reform by High School GPA

	High School GPA								
	Low (≤ 2.8)			Average (> 2.8 & ≤ 3.2)			High (> 3.2)		
	Blank	Test Score	Failure	Blank	Test Score	Failure	Blank	Test Score	Failure
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Reform 2008	0.029	1.870	-0.048	0.079***	-7.465*	0.046	0.036*	-1.600	-0.014
	(0.029)	(4.227)	(0.108)	(0.027)	(4.201)	(0.055)	(0.019)	(4.641)	(0.064)
2010	0.010	-18.826***	0.113	-0.005	-14.521***	0.004	0.003	-18.533***	0.059
	(0.011)	(2.763)	(0.070)	(0.006)	(3.768)	(0.043)	(0.004)	(3.970)	(0.053)
Bus. Admin.	0.015	-4.792*	0.152**	-0.002	-2.361	0.084**	0.013	-2.912	0.051
	(0.016)	(2.720)	(0.063)	(0.012)	(2.482)	(0.033)	(0.011)	(3.322)	(0.038)
Male	-0.001	3.981*	-0.046	0.005	4.560**	-0.088**	0.006	-3.402	0.029
	(0.007)	(2.190)	(0.055)	(0.021)	(1.813)	(0.034)	(0.019)	(2.164)	(0.034)
Age	-0.002	-1.004**	0.018**	0.003	-0.930	0.023*	0.003	-0.472	0.010
	(0.001)	(0.476)	(0.009)	(0.005)	(0.608)	(0.014)	(0.003)	(0.633)	(0.009)
Foreign	0.288**	1.799	0.147	0.023	-14.188***	0.274***	-0.030*	-12.134***	0.142*
	(0.140)	(5.973)	(0.177)	(0.051)	(4.846)	(0.097)	(0.017)	(3.872)	(0.075)
Semester	0.008	-5.207**	0.188***	0.015	6.823*	-0.147*	-0.008	6.057**	-0.020
	(0.006)	(2.595)	(0.056)	(0.012)	(3.677)	(0.080)	(0.007)	(3.013)	(0.023)
High School GPA	-0.001	6.286**	-0.069	-0.045	17.216**	-0.193	0.009	5.013	0.018
	(0.037)	(3.184)	(0.086)	(0.087)	(8.590)	(0.156)	(0.064)	(5.983)	(0.082)
Blank		-52.537***	0.590***		-58.413***	0.866***		-67.732***	0.925***
		(5.331)	(0.142)		(1.918)	(0.038)		(2.411)	(0.033)
Constant	0.021	98.079***	-0.729**	0.027	31.170	0.528	-0.083	68.753**	-0.264
	(0.101)	(15.030)	(0.357)	(0.243)	(31.114)	(0.668)	(0.253)	(27.350)	(0.355)
Observations	264	264	264	288	288	288	216	216	216
R^2	0.236	0.388	0.136	0.044	0.577	0.354	0.027	0.503	0.257

Notes: This table reports OLS estimates of the effect of the 2008 reform on the rate of blank submissions, test scores, and the rate of failure by high school GPA. The variable “Reform 2008” is an interaction of the dummy variable for the year 2010 and the business administration program dummy variable. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

dubious whether students who fail their first three attempts are more likely to succeed at their forth or fifth attempts or whether they drop out in their final semesters.

Figure 4.3. Average Test Score and Rate of Failure by Number of Attempts.



Notes: This figure shows the average test score and rate of failure by the number of attempts. Students that submitted blank are not considered for this figure since those artificially decrease average test scores and increase the average rate of failure.

Unfortunately, we were unable to gather data on graduations and final grades of the students in our data. However, Figure 4.3 shows that students at a higher attempt scored in general lower and were more likely to fail. Obviously, there is adverse selection because students must not resit the exam once they have passed and consequently only those students who performed worst remain for the next attempt. The effects are therefore not only driven by the additional workload and strain due to procrastination and differences in motivation and effort, but also by differences in ability. Nevertheless, this figure questions whether the higher number of resits does help students to succeed.

4.7 Robustness Tests

Our study exploits the fact that two very similar study programs have been reformed at different points of time. So far, our results suggest that the reform of 2005 had a negative effect on student test scores and that the reform of 2008 had a positive effect on blank submissions. We now test the robustness of our results.

4.7.1 Consistency Check

In order to isolate the effects of the 2005 reform (that came into effect between 2006 and 2008) and the 2008 reform (that came into effect between 2008 and 2010), we compared business administration students with business education students between 2006 and 2008 and 2008 and 2010, respectively.

To check for consistency of the results, we analyze the effect of the revised Bachelor program by comparing the data of 2006 and 2010. In 2006, both programs studied under the *Diplom* program. In 2010, business education students were still studying under the *Diplom*, but business administration students were enrolled in the revised Bachelor program which involves both the Bachelor certification instead of the *Vordiplom* (reform 2005) and the higher number of allowed resits (reform 2008). Consequently, we need to account for the fact that these students were allowed to resit exams more often when analyzing the effect on their test scores and failure rates. Table 4.7 confirms the previous results. In particular, we find a negative effect on test scores which is not driven by blank submissions (column 3). The coefficients are a bit larger as those reported in Table 4.2. Furthermore, a positive effect on blank submission is reported which is about the same size as reported in Table 4.4.

The same pattern is also reported in a regression that pools the data of 2006, 2008, and 2010 and includes interaction variables for 2008 and the program of business administration (2008xBA) and 2010 and the program of business administration (2010xBA). Table 4.8 shows the negative effect on test scores between 2006 and 2008 (2008xBA), which refers to the comparison of business administration students enrolled in the *Diplom* program and business administration students enrolled in the Bachelor program to business education students. The negative effect on test scores is also reported between 2006 and 2010 (2010xBA), which refers to the comparison of business administration students enrolled in the *Diplom* program and business administration students enrolled in the *revised* Bachelor program to business education students. Furthermore, we find the positive effect on the rate of blank submission only between 2006 and 2010 (2010xBA), which refers to the comparison of business administration students enrolled in the *Diplom* program and business administration students enrolled in the *revised* Bachelor program to business education students, but not between 2006 and 2008 (2008xBA).

Table 4.7
Robustness: Effects of the 2005 and 2008 Reforms using data of 2006 and 2010

	Test Score			Failure			Blank		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Reforms	-11.658***	-14.082***	-10.462***	0.122	0.155*	0.070	0.056***	0.055***	0.056***
	(3.593)	(3.271)	(3.251)	(0.094)	(0.089)	(0.089)	(0.014)	(0.014)	(0.014)
2010	-2.229	-2.411	-2.808	-0.161*	-0.159**	-0.130*	-0.000***	-0.001	-0.002
	(2.838)	(2.533)	(2.584)	(0.083)	(0.079)	(0.078)	(0.000)	(0.002)	(0.003)
Bus. Admin.	9.825***	7.612***	6.780**	-0.101	-0.072	-0.006	-0.000***	-0.000	-0.001
	(2.939)	(2.658)	(2.825)	(0.084)	(0.080)	(0.082)	(0.000)	(0.004)	(0.007)
Male		1.438	1.396		-0.015	-0.022		-0.003	-0.003
		(1.658)	(1.431)		(0.034)	(0.031)		(0.014)	(0.014)
Foreign		-10.723***	-7.223**		0.191**	0.142**		0.063	0.063
		(4.009)	(3.447)		(0.074)	(0.069)		(0.049)	(0.049)
High School GPA		13.466***	12.495***		-0.194***	-0.151***		-0.010	-0.010
		(1.625)	(1.564)		(0.033)	(0.033)		(0.013)	(0.014)
Age			-0.347			0.018**			-0.000
			(0.331)			(0.009)			(0.002)
Semester			0.287			0.064*			0.001
			(1.330)			(0.034)			(0.003)
Blank			-57.998***			0.824***			
			(2.597)			(0.038)			
Constant	62.590***	26.098***	36.625***	0.333***	0.856***	0.074	0.000***	0.027	0.037
	(2.342)	(5.039)	(11.773)	(0.076)	(0.122)	(0.293)	(0.000)	(0.037)	(0.079)
Observations	575	575	575	575	575	575	575	575	575
R^2	0.070	0.188	0.388	0.010	0.079	0.202	0.029	0.039	0.039

Notes: This table reports OLS estimates of the effect of the “Revised Bachelor Program”, that comprises the elements of the 2005 and 2008 reform, on test scores, the rate of failure, and the rate of blank submission. The variable “Reforms” is an interaction of the dummy variable for the year 2010 and the business administration program dummy variable. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 4.8

Robustness: Effects of the 2005 and 2008 Reforms using data of 2006, 2008, and 2010

	Test Score			Failure			Blank		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2008 x BA	-9.751***	-9.628***	-8.476***	0.144	0.141	0.102	0.003	0.003	0.003
	(3.555)	(3.105)	(3.112)	(0.092)	(0.087)	(0.086)	(0.003)	(0.004)	(0.004)
2010 x BA	-11.658***	-14.268***	-9.923***	0.122	0.155*	0.068	0.056***	0.057***	0.057***
	(3.592)	(3.266)	(3.226)	(0.094)	(0.089)	(0.088)	(0.014)	(0.014)	(0.014)
2008	17.109***	15.801***	15.009***	-0.251***	-0.234***	-0.207***	0.000***	0.000	-0.000
	(2.883)	(2.505)	(2.523)	(0.082)	(0.077)	(0.076)	(0.000)	(0.002)	(0.002)
2010	-2.229	-2.415	-3.273	-0.161*	-0.160**	-0.129*	0.000***	-0.001	-0.001
	(2.838)	(2.544)	(2.581)	(0.083)	(0.078)	(0.078)	(0.000)	(0.002)	(0.002)
Bus. Admin.	9.825***	7.509***	5.479**	-0.101	-0.069	0.002	0.000	0.001	-0.000
	(2.938)	(2.649)	(2.760)	(0.084)	(0.079)	(0.080)	.	(0.003)	(0.004)
Male		1.238	1.505		-0.028	-0.036		-0.000	-0.000
		(1.192)	(1.071)		(0.024)	(0.023)		(0.009)	(0.009)
Foreign		-12.647***	-10.425***		0.214***	0.177***		0.037	0.037
		(2.872)	(2.639)		(0.057)	(0.055)		(0.029)	(0.029)
High School GPA		14.709***	12.903***		-0.201***	-0.154***		-0.011	-0.012
		(1.293)	(1.212)		(0.026)	(0.025)		(0.010)	(0.011)
Semester			0.501			0.057*			-0.000
			(1.356)			(0.033)			(0.002)
Age			-0.851***			0.021***			-0.000
			(0.297)			(0.006)			(0.001)
Blank			-57.416***			0.808***			
			(2.718)			(0.040)			
Constant	62.590***	22.840***	47.657***	0.333***	0.881***	0.039	-0.000***	0.030	0.039
	(2.342)	(4.209)	(9.869)	(0.076)	(0.105)	(0.216)	(0.000)	(0.027)	(0.052)
Observations	936	936	936	936	936	936	936	936	936
R^2	0.197	0.326	0.456	0.021	0.106	0.205	0.036	0.042	0.042

Notes: This table reports OLS estimates of the effect of the 2005 and 2008 reform on test scores, the rate of failure, and blank submissions. The variable "2008xBA" is an interaction of the dummy variable for the year 2008 and the dummy variable of the business administration program. The variable "2010xBA" is an interaction for the dummy variable of the year 2010 and the dummy variable of the business administration program. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

The difference between the coefficients on the interaction variables is the reform effect between 2008 and 2010. As the 2005 reform came already into effect between 2006 and 2008, we find no difference on test scores between 2008 and 2010, which is reflected in the similar coefficients of the interaction variables. In contrast, as the increase in allowed resits came into effect between 2008 and 2010, the difference between the coefficients of the interaction variables reflects the reform effect of about the same size between 2008 and 2010.

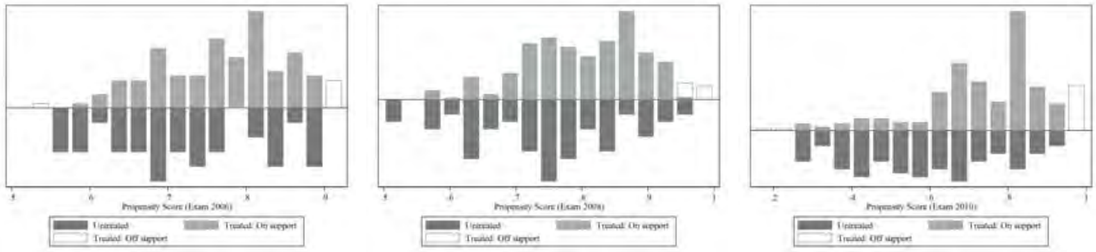
4.7.2 Specification Tests

As can be seen from Table 4.1, there are differences in age, semesters and high school GPAs between the two groups and over time. Interestingly, the main results in Tables 4.2 and 4.4 are not sensitive to the inclusion of these control variables which confirms that the treatment effect is not driven by these differences. Nevertheless, we perform our analyses on different subsamples that should be even more comparable. First, we address the fact that business education students are on average older by restricting the analysis to students in the 90^{th} percentile of the age distribution which refers to an age of 26 years. Second, we restrict our sample to students in their third semesters. As the exam is recommended to be taken in the third semester, this group does not include students who procrastinated the exam or who take the exam earlier in their studies for some reasons.

Finally, we perform propensity matching methods to compare the treated individuals to the most similar non-treated individuals in terms of observable characteristics. In particular, we estimate the propensity score based on the control variables that are predetermined before students enroll at university, namely gender, migrant status and high school GPA (see Figure 4.4).

We then perform two matching techniques, namely nearest neighbor caliper matching with replacement and kernel matching. For both techniques, we impose a common support which means that we drop treated individuals that have an estimated propensity score less than the minimum or more than the maximum of the controls. For the nearest neighbor matching, we choose a matching partner for each treated student that is closest in terms of the estimated propensity score and lies within the caliper of 0.04 (which

Figure 4.4. Estimated Propensity Scores by Year of Examination.



Notes: The figure reports estimated propensity scores by year of examination after performing kernel matching by imposing a common support. Variables used to estimate the propensity score are gender, migrant status and high school GPA.

refers to 0.2 of the standard deviation of the estimated propensity score).[8] We then run our difference-in-differences regression on this sample. Since some of the non-treated observations are matched several times in the replacement case, we weight the regression by the frequencies with which the non-treated observation are used as matches. We additionally perform kernel matching, where non-treated observations that are more similar in terms of the estimated propensity score receive a higher weight.

Table 4.9 reports the effect of the 2005 reform on test scores and the effect of the 2008 reform on blank submissions in the two subsamples along with the matching results. The left-hand side of this table shows a negative, significant effect of the 2005 reform on test scores. While the effect in the group of third semester students is about one standard deviation larger, the coefficients in columns (2) to (4) are similar to those reported in Table 4.2. The right-hand side of this table shows a positive, significant effect on the rate of blank submission. The size and significance of the coefficients correspond to those reported in Table 4.4. Although we report the results while controlling for student characteristics, the results are not sensitive to the inclusion of control variables. As in Table 4.2, the effect of the 2005 reform on the rate of failure is positive, but not significant (not shown). Overall, this table confirms the previous results.

4.7.3 Placebo Test

We argued above that the common trend assumption is likely to be satisfied because of the unique setting that we consider. Both groups of students attend the same course, are

[8] Smith and Todd (2005) note that it is difficult to know a reasonable caliper width a priori. We follow Cochran and Rubin (1973) and use a caliper width of .2 standard deviations.

Table 4.9
Robustness: Main Results for different Subsamples and Matching Results

	Reform 2005				Reform 2008			
	Test Score				Blank			
	3. Semester[a]	Age[b] $\leq$ 26	Near. Neigh.[c]	Kernel[d]	3. Semester[a]	Age[b] $\leq$ 26	Near. Neigh.[c]	Kernel [d]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform	-14.535***	-9.485***	-7.784***	-11.132***	0.056***	0.061***	0.049***	0.052***
	(3.915)	(3.374)	(2.233)	(3.196)	(0.014)	(0.015)	(0.014)	(0.015)
Year	17.851***	16.346***	13.957***	16.730***	-0.001	-0.003	0.004	0.000
	(3.291)	(2.828)	(1.264)	(2.553)	(0.002)	(0.003)	(0.002)	(0.003)
Bus. Admin.	10.773***	8.380***	3.531*	5.927**	0.003	0.005	0.003	0.003
	(3.751)	(2.988)	(1.978)	(2.866)	(0.006)	(0.006)	(0.004)	(0.005)
Male	0.269	0.977	0.093	2.277	0.001	-0.000	-0.000	-0.001
	(1.424)	(1.389)	(1.023)	(1.385)	(0.011)	(0.012)	(0.007)	(0.007)
Foreign	-15.193***	-17.217***	-17.485***	-8.944	0.042	0.048	0.075	0.056
	(3.752)	(4.013)	(3.779)	(6.395)	(0.033)	(0.036)	(0.047)	(0.038)
High School GPA	14.350***	16.242***	13.768***	15.228***	-0.016	-0.024	-0.005	-0.008
	(1.730)	(1.631)	(1.158)	(1.670)	(0.015)	(0.016)	(0.007)	(0.007)
Age	-1.142***		-1.573***	-1.313***	-0.001		-0.000	-0.001
	(0.391)		(0.275)	(0.347)	(0.002)		(0.001)	(0.001)
Semester		1.366	1.718	1.034		0.003	0.001	0.001
		(1.643)	(1.115)	(1.500)		(0.006)	(0.002)	(0.003)
Constant	50.623***	13.895*	61.173***	49.672***	0.060	0.058	0.019	0.034
	(12.303)	(7.695)	(7.691)	(11.376)	(0.067)	(0.057)	(0.032)	(0.035)
Observations	444	493	794	502	736	712	1056	723
R^2	0.274	0.300	0.389	0.385	0.039	0.044	0.051	0.049
Data	2006 and 2008				2008 and 2010			

Notes: The left-hand side of this table reports OLS estimates of the effect of the 2005 reform on test scores. The right-hand side of this table reports OLS estimates of the effect of the 2008 reform on blank submissions. The variable “Reform ” is an interaction of the dummy variable of the year 2008 (2010, respectively) and the dummy variable of the business administration program. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.
[a]Sample includes students in their third semesters only.
[b]Sample includes students $\leq$ 26 years.
[c]Sample refers to treated observations and their nearest neighbors within a caliper of 0.04 while imposing a common support. Non-treated observations are weighted by the frequency they are used as a match.
[d]Sample includes all treated and all non-treated observations while imposing a common support. Non-treated observations are weighted in proportion to their proximity to treated observations in terms of estimated propensity scores.

taught by the same lecturer, write the same exam and so forth. Ideally, the common trend assumption is tested by using other pre-treatment time periods to verify that there are no pre-existing differences in trends. Unfortunately, we were unable to obtain sufficient data from 2004 or earlier years because the office of the university registrar did not keep records at that time. However, we try to exploit the fact that the business education program was reformed in one single step in 2010, which comprised both reforms of the business administration program. Thus, we consider the years of 2006 and 2012 and check whether the two groups follow the same trend given both of them have been reformed in the same way (see Figure 4.2). Unfortunately, students who enrolled in business education in 2012 received the recommendation to take the exam considered here in their first semesters. Thus, the universe of business education students in 2012 is generally younger and takes the exam on average in lower semesters . Nevertheless, we compare the development of business administration students between 2006 and 2012 against the development of business education students while controlling for age and semester. Since both groups of students were treated identically between 2006 and 2012, however at different points of time, we should not find a significant treatment effect.

For blank submissions, Table 4.10 shows a coefficient that is close to zero and statistically not significant. This suggests that both groups of students follow the same trend before and after the increase in the number of resits. For test scores, however, we see a negative, significant effect in columns (1) and (2). Consequently, the effect on the rate of failure is positive, but not significant. However, as can be seen from Table 4.1, there is a large difference in students' age in 2006 which has almost completely vanished in 2012. Conversely, there is no difference in students' semesters in 2006, but in 2012. After conditioning on students' age and semesters, the coefficient decreases strongly in magnitude and significance. In other words, conditional on students' age and semesters, we find no significant difference in students' responses to the reforms.

While the negative effect of the 2005 reform on test scores of business administration students has been confirmed in several specifications, the results of the placebo test without controlling for age and semester suggest that the negative effect on test scores may only be temporary. This could by explained by confusion and uncertainty among students due to the introduction of the comprehensive Bachelor reform in 2005. However, since students used to take the exam in their third semesters, they were already quite familiar with the

Table 4.10
Robustness: Placebo Test using data of 2006 and 2012

	Test Score			Rate of Failure			Blank Submission		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2012 x BA	-8.819**	-7.373**	-2.092	0.096	0.076	-0.008	-0.000	-0.002	-0.028
	(3.856)	(3.477)	(3.669)	(0.093)	(0.088)	(0.089)	(0.019)	(0.020)	(0.022)
Bus. Admin.	9.825***	8.008***	4.249	-0.101	-0.077	-0.019	0.000***	0.002	0.014
	(2.937)	(2.651)	(2.827)	(0.084)	(0.079)	(0.082)	(0.000)	(0.003)	(0.011)
2012	16.273***	12.663***	8.313***	-0.190**	-0.145*	-0.077	0.031*	0.035*	0.056**
	(3.237)	(2.937)	(3.146)	(0.084)	(0.079)	(0.080)	(0.017)	(0.018)	(0.022)
Male		-1.851	-1.434		0.005	-0.002		0.005	0.003
		(1.597)	(1.584)		(0.028)	(0.028)		(0.012)	(0.011)
Foreign		-15.833***	-15.410***		0.296***	0.289***		0.002	-0.001
		(4.288)	(4.158)		(0.074)	(0.071)		(0.026)	(0.025)
High School GPA		15.986***	13.799***		-0.209***	-0.175***		-0.019	-0.012
		(1.518)	(1.546)		(0.031)	(0.032)		(0.012)	(0.011)
Age			-1.349***			0.021***			0.003
			(0.339)			(0.007)			(0.003)
Semester			-0.962			0.017			0.014
			(0.934)			(0.018)			(0.010)
Constant	62.590***	20.454***	63.704***	0.333***	0.891***	0.224	-0.000***	0.049	-0.085
	(2.341)	(4.802)	(11.144)	(0.076)	(0.118)	(0.225)	(0.000)	(0.033)	(0.098)
Observations	694	694	694	694	694	694	694	694	694
R^2	0.042	0.180	0.202	0.021	0.120	0.140	0.008	0.012	0.021

Notes: This table reports OLS estimates on test scores, the rate of failure, and the rate of blank submission. The variable “2012xBA” is an interaction of the dummy variable for the year 2012 and the dummy variable of the business administration program. Robust standard errors appear in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

program and its policies. Moreover, the fact that we find the negative effect on test scores also as a result of the revised Bachelor program, which has only been slightly modified, points to more than a restructuring effect.

4.8 Conclusions

The analyses presented in this chapter suggest that university and program policies serve as incentives. By using a difference-in-differences approach, we analyze the effect of modified program policies on student test scores, failure rates, and blank submissions. Our results indicate that the 2005 reform, which effectively doubled the time until students receive their first certificate and which reduced the impact of each exam on the GPA, has a significant, negative impact on students' test scores. While there is a positive effect on the rate of failure, the coefficient is statistically not significant. The 2008 reform, which increased the number of allowed resits, increases the share of students that submitted blank papers. Since students must not resit exams once they have passed, they can resit the exam and improve by failing the exam deliberately. On average, however, students do not perform worse if they have the opportunity to resit exams as often as they want.

Moreover, we find that students respond differently to the reforms depending on their ability. In particular, we find that poor students do not react to the 2005 reform which indicates that those are not able to influence their performance via their efforts. In contrast, for average or excellent students we observe a decrease in test scores which might result from lower effort choices. This is in line with the finding that only high ability students (are able to) respond to financial incentives (Leuven et al., 2010). Furthermore, we do not find a positive effect on blank submissions among the group of poor performing students. This is also intuitive since poor students hope that they may have just passed or done sufficiently well rather than deliberately failing. Overall, this suggests that poor students maximize a different objective function, which needs to be considered in designing incentives for students.

Our identification relies on the fact that we observe two very similar study programs that were reformed at different points of time. Important to our quasi-experimental approach is the fact that both groups of students have a nearly identical curriculum and are required to take the same exam which is taught by the same lecturer in all periods. Moreover, the placebo test suggests that both groups follow the same trend after conditioning on students' age and semesters. At the same time, it is appropriate to close with caution. Although the results are robust to a number of specifications and tests, there is some evidence that the negative effects on test scores are only temporary.

Nevertheless, this analysis introduces a new perspective on university and program policies and suggests to consider them as incentives. Since policy choices on the grading system, the level of fees, the duration of programs and the number of courses and resits are necessary for the design of study programs, it is important to understand how they influence student effort choices. Moreover, these policies are a promising field of research, because they are available to every university and, in contrast to monetary rewards, generally inexpensive.

Chapter 5

Forced Migration and the Effects of an Integration Policy in Post-World War II Germany*

5.1 Introduction

We study the effect of an integration policy in the context of a forced mass migration that occurred in the aftermath of World War II. Significant territorial changes forced 8 million of ethnic Germans, hereafter expellees, to leave their homelands in East Prussia, Silesia, Pomerania, and Bohemia and settle within the new borders of West Germany (cf. Schmidt, 1994). This was possibly one of the largest mass migration shocks ever experienced by a developed country in modern history. After their displacement, many expellees experienced a huge loss in status. While many of them owned real estate or were self-employed before WWII, large fractions of expellees became occupied in low skilled jobs or even unemployed.

After the Federal Republic of Germany was founded in 1949, the Federal Expellee Law (Bundesvertriebenengesetz) was introduced in 1953 as a reaction to the expellees' overall bad economic situation. The law aimed at improving the economic situation of the expellees. For this purpose, the law instructed public employment services to

* This chapter was coauthored by Oliver Falck and Stephan Heblich and was published as "Forced Migration and the Effects of an Integration Policy in Post-World War II Germany", *B.E. Journal of Economic Analysis & Policy: Topics* 12 (1) 2012, De Gruyter.

consider expellees first as long as local unemployment rates among expellees were higher than among local West Germans. Public employment services were further instructed to assist expellees in finding a job equivalent to their occupation prior to WWII and to promote self-employment and entrepreneurship in agriculture and non-agricultural sectors. Incentives for self-employment and entrepreneurship included credits at subsidized interest rates or tax credits. Moreover, short-run credits were converted into long-run credits at subsidized interest rates to support expellees who were already entrepreneurs or self-employed. Taken together, this leaves us with the following three aspects targeted by the Federal Expellee Law: (i) Transition from unemployment to employment; (ii) restitution of previous or comparable occupations in case of degradation; (iii) promotion of entrepreneurship and self-employment.

To evaluate whether the Federal Expellee Law met these goals, we exploit data from the 1971 micro census that allow us to identify and distinguish expellees from local West Germans.[1] Furthermore, the 1971 census contains a special survey that provides retrospective information about the respondents' occupation in 1939, 1950, 1960 and 1971. Put differently, we have retrospective information about the occupational status of individuals at one point before WWII, one point after WWII but before the introduction of the law, and two points after the introduction of the law. Based on this information, we create a longitudinal dataset for the period 1939 to 1971. Depending on the outcome targeted by the law, we define different comparison groups drawn from the population of local West Germans and, in the case of entrepreneurship and self-employment, from the population of refugees from the GDR.[2] These refugees constitute an interesting comparison group because, similarly to the expellees, they lost their property and social contacts due to their flight. However, apart from a small group of political refugees, GDR refugees did not benefit from the law.

Our empirical analyses compare the observed occupational status of expellees and local West Germans in the years 1960 and 1971 conditional on socio-demographic characteristics. To disentangle the integration effect of the Federal Expellee law from a more general catch-up process of expellees in times of dramatic economic growth where unemployment among local West Germans was close to zero, we match expellees and local West Germans

[1] For a description of these data in the context of expellees, see Lüttinger (1989).

[2] Before 1949, the GDR was the Soviet zone of occupation in Germany. For simplicity, we will refer to it as GDR throughout the paper.

based on their economic situation in 1950.[3] This approach should uncover even small effects of the law since local West Germans found themselves in a better economic situation in 1950 than expellees with similar socio-demographic characteristics. One indication is for instance that most West Germans worked in an occupation similar to the one before WWII in 1950. Those West Germans who faced – similar to the expellees – a worse economic situation in 1950 than before WWII were apparently a selective group although they resemble the expellees in their characteristics. A comparison between this group of local West Germans and the expellees may thus favour a positive effect of the law. Despite the possibility of an upward bias in our estimations, we find no evidence that the law met its defined goal to foster expellees' labor market integration.

Our research contributes to other recent papers that employ micro-level data to analyze the assimilation of individuals who were expelled from their homelands following territorial changes in the aftermath of WWII (cf. Braun et al., 2012; Sarvimäki et al., 2009, for Germany and Finland). Specifically, we are interested in the success of an integration policy in Germany and attempt to disentangle general assimilation effects. One may question the external validity of our results arguing that this historical episode is a period of dramatic growth leaving the Federal Expellee Law with no comparable successors. By contrast, we argue that parts of the law indeed have their contemporaneous counterparts. Many active labor market policies today put emphasis on public employment services to assist unemployed workers in finding jobs and entrepreneurship policies try to alleviate financing constraints of potential entrepreneurs by granting credits at reduced interest rates and tax credits. In the meantime, evaluation studies have provided a lot of evidence on the effectiveness of public employment services (for an overview, cf. Heckman et al., 1999). However, our knowledge about the success of entrepreneurship policies is still limited. This chapter is thus not only a piece in economic history but also contributes to other literature strands, among others on the effects of entrepreneurship polices.

The remainder of this chapter is organized as follows. Section 5.2 provides the historical background. Section 5.3 introduces our data and provides descriptive statistics on the socio-demographic characteristics of expellees and their development in the labor market.

---

[3] There is a huge literature on the assimilation of immigrants. For an overview see, e.g., Borjas (1994, 1999); Pekkala Kerr and Kerr (2011).

In Section 5.4, we present estimation results on the impacts of the Federal Expellee Law on various labor market outcomes targeted by the law. Section 5.5 concludes.

5.2 Historic Context

The significant territorial changes that occurred in the aftermath of WWII resulted in large migration streams across Europe. The biggest of these involved almost 8 million ethnic Germans who were forced by the Red Army and, after WWII, the Potsdam Treaty to leave their homelands, predominantly East Prussia, Pomerania, Silesia, East Brandenburg, and the Sudetenland, and settle within the new borders of West Germany. This forced mass migration affected all ethnic German individuals regardless of their social status or skill level (Bethlehem, 1982; Schmidt, 1994). Table 5.1 illustrates the distribution of expellees across West German states in absolute numbers, as a fraction of the expellee population, and as a fraction of the local West German population.

The mechanism for allocating expellees across settlement states worked as follows. In the period between the end of WWII in 1945 and the founding of the two separate German states in 1949, the allied powers divided Germany into four occupation zones. Figure 5.1 shows the four occupation zones along with the predominantly ethnic German areas where the expellees lived before WWII. In 1949, the French, British, and U.S. zones of occupation were merged into the Federal Republic of Germany (FRG) and the Soviet zone became the socialist German Democratic Republic (GDR). Table 5.1 reveals that there was an especially pronounced difference in the number of expellees in the French occupation zone compared to the other zones. This is due to the French authorities' desire to limit the number of people competing for already scarce resources (Grosser, 2001). As a result, initially Rhineland-Palatine and the French-occupied areas in Baden-Württemberg received no expellees.[4] Authorities in the other zones distributed the expellees according to a central formula based on the availability of nutrition and housing space. Since most German cities were destroyed and nutrition and housing were more plentiful in rural areas, the vast majority of expellees were settled in the countryside (cf. Brakman et al., 2004; Grosser, 2006).

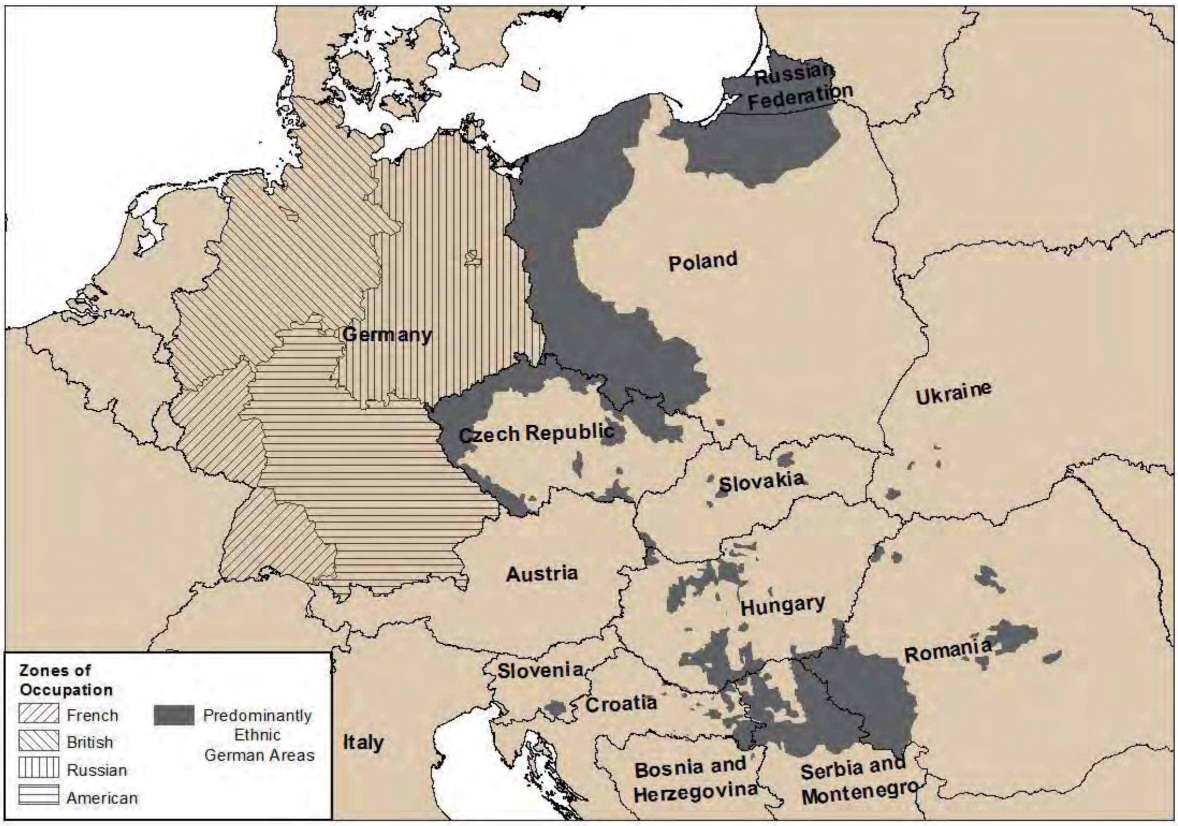
[4] Only after 1948 did authorities reallocate expellees from regions with a high number to regions with fewer, particularly to the French regions. Thus, by 1956, about 1 million expellees had been forced to relocate again.

Table 5.1
Expellees by State in Post-War Germany in 1950

State	Occupation Zone	Number of Expellees	% of Expellees	% of State Population
Bavaria	A	1,937,000	16.2	21
Lower Saxony	B	1,851,000	15.5	27
North Rhine-Westphalia	B	1,332,000	11.2	10
Baden-Wuerttemberg	F/A	862,000	7.2	13.5
Schleswig-Holstein	B	857,000	7.2	33
Hessen	A	721,000	6	16.5
Rhineland-Palatinate	F	152,000	1.3	5
West-Berlin	A/F/B	148,000	1.2	7
Hamburg	B	116,000	1	7
Bremen	A	48,000	0.4	9
West Germany	A, B and F	8,024,000	67.2	15.75

Notes: A = American, B = British, F = French, S = Soviet. Source: Federal Statistical Office.

Figure 5.1. Zones of Occupation and Predominantly Ethnic German Areas.



Notes: This figure shows the zones of occupation and the predominantly ethnic German areas.

Furthermore, free movement within the territory of the German Federal Republic was restricted by Allied legislation until June 1950.[5] The economic situation of most expellees was precarious. Some of them were able to meet at least a part of their needs by working as unskilled labor in the agricultural sector, but many suffered hunger and had to beg or steal to fulfill their basic needs (Vaskovics, 2002). In many regions, expellees were viewed as a burden and this was reflected in governmental restrictions on their rental contracts (Schraut, 1995). Often, expellees were refused the permits necessary for starting a business. Attaining recognition of formal occupational qualifications, e.g., certificates for lawyers or doctors, was complicated (Müller, 1993; Schraut, 1995). There were barriers to accessing capital because banks did usually not provide credit to expellees without any

[5] Unfortunately, the micro census does not provide information on the migration process of expellees after they were allocated across West Germany. However, on the basis of regional-level data from the population censuses in 1950 and 1961, we calculated the correlation coefficient of the share of expellees in 1950 and 1961 across regions. It is about 0.82 and highly significant. We conclude by this that the mobility of expellees after 1950 was rather limited.

collateral. As a result, in 1950, only a small fraction of expellees worked in the same field or occupation as they had in 1939 (Schraut, 1995).

In 1953, the German government introduced the Federal Expellee Law (Bundesvertriebenengesetz) with the goal of restoring the expellees' status quo and improving their situation.[6] The law provided official acceptance and legitimation for a wide range of occupational certificates held by expellees, including those of doctors, dentists, and craftsmen (§§69-71). The law instructed public employment services to first place expellees as long as local unemployment rates among expellees were higher than among local West Germans. Public employment services were further instructed to assist expellees in finding a job equivalent to their occupation prior to WWII (§§77-79). The law improved access to start-up capital and provided tax incentives for self-employment and entrepreneurship. Converting short-run credits to entrepreneurs and self-employed individuals into long-run credits at a subsidized interest rate was further designed to reduce exits among expellees who were already entrepreneurs or self-employed (§§72 and 73). The law offered better opportunities to rent state-owned property for business purposes (§76), and ensured that businesses run by expellees were treated preferentially when public contracts were awarded (§§74 and 75). Finally, the law also helped integrate those who had been farmers prior to WWII into the agriculture sector (§§35-68).

The eligibility requirements to benefit from these privileges were tied to the official status as an expellee (Categories A and B). This status was defined in Section 1 of the 1953 Federal Expellee Law, and defines an expellee as being either a German citizen or an ethnic German who before and/or during WWII lived within the 1917-1937 borders of eastern Germany and Austria-Hungary.[7] In addition to expellees, political refugees from the socialist GDR (and, prior to 1949, the Soviet zone) were also covered by this law (§3). However, to qualify as "eligible refugees" (Category C), GDR refugees had to prove that they had suffered "a direct threat to life and limb or their personal freedom" (Ackermann, 1995, p.13).

[6] From 1949-1969, the Federal Republic of Germany formed the "Federal Ministry for Displaced Persons, Refugees and War Victims" that was part of the West German government. It was responsible to coordinate the integration of displaced persons and refugees, care for war victims and provide compensation and initial aid. This ministry also enacted the Federal Expellee Law in 1953.

[7] Since we employ a twofold definition of expellees based on their residence in 1939 and the possession of an expellee pass, we count expellees who migrated to the GDR first and then to West Germany as expellees and not as GDR refugees.

In the empirical section of this paper, we focus on expellees in Categories A and B only, omitting from our analysis those GDR refugees (Category C) who were covered by the Federal Expellee Law. Given their political motives for leaving East Germany, GDR refugees are probably a highly distinctive group and including them in our empirical analyses could bias our estimates of the effect of the Federal Expellee Law. However, in some specifications, we will use GDR refugees who are not covered by the law as a comparison group. The GDR refugees who were not covered by the Category C of the Federal Expellee Law were looking for political freedom and economic prosperity (cf. Ackermann, 1995; Heidemeyer, 1994; Hoffmann, 2000).[8] Altogether, more than 2.75 million people fled East Germany to resettle in West Germany prior to the construction of the Berlin Wall in 1961 and, like the expellees, the refugees from the GDR were at first centrally distributed across the federal states according to §17(1) of the 1950 provisional accommodation law (Notaufnahmegesetz). The provisional accommodation law granted some financial support to the GDR refugees, but it was far less extensive than that available under the Federal Expellee Law.

5.3 Descriptive Statistics between 1939-1971

Our data are drawn from the German micro census 1971. The micro census consists of a one percent sample of the German population and provides representative cross-sectional statistics for the population and labor market in Germany.[9] The micro census of 1971 includes an extension (MZU, 1971) that was designed to gain insight into expellees' integration into the German labor market and society. This extension is particularly interesting for our analysis because it contains detailed retrospective information on the occupation of the German population in 1939, 1950, and 1960 as well as the place of

[8] As it became clear that Germany's separation was permanent and that East Germany was adopting a Soviet system a first wave of GDR refugees included a large number of civil servants. In our sample, about seven percent of GDR refugees worked in the civil service before WWII compared to four to five percent of West Germans and expellees. This is because the Soviet authorities abolished the civil service system and because the denazification was more rigorous in the GDR leaving more people who were public employees during the Nazi era without a job. Only later when it became apparent that collectivization of agriculture was imminent, did individuals working in this sector leave East Germany.

[9] The micro census is a random sample combining a one-stage cluster sample design with a partial rotation procedure. In each sampling district, chosen from within the territory of the Federal Republic of Germany, all households and persons are interviewed. Every year, one-quarter of the sample households is replaced.

residence in 1939, home ownership in 1939, and the year of arrival within the new borders of West Germany.[10] Our analyses concentrate on individuals who have finished education and transited into the labor market by 1939. All socio-demographic characteristics of the respondents stem from the year 1971. Since our sample is restricted to individuals who have finished education before 1939, the variable highest educational degree obtained is time-invariant. Other characteristics like age at arrival are counted back from the 1971 information. We end up with retrospective longitudinal data ranging from 1939 to 1971. This leaves us with individual information from one point in time before displacement (1939), one point in time after displacement but before the Federal Expellee Law set in (1950) and two points in time after the law was introduced (1960 and 1971).

Our definition of expellee status is twofold. We define a person as an expellee if he or she (i) possesses a Category A or B pass and (ii) lived in the former eastern territories of the German Reich or Austria-Hungary in 1939. We only consider expellees who arrived within the new borders of Germany between 1945 and 1950 because they were forced to migrate immediately after WWII. We drop individuals who came during the Nazi regime or who voluntarily arrived after 1950 in search of economic opportunities. After excluding individuals with missing data on occupational status, our sample contains 23,183 expellees. The sample further includes 146,786 local West Germans and 2,826 GDR refugees who migrated to West Germany between 1945 and 1950, 1,896 of whom were not accepted as political refugees and not covered by the Federal Expellee Law. Our final sample consists of about 13.5 percent expellees and 1.6 percent refugees from the Soviet zone of occupation. Given an overall population of roughly 50 million in West Germany in 1950, this sample is a good representation of the population shares, i.e., the group of expellees (8 million) being about 15 percent of the West German population and the refugees from East Germany (2.75 million until the construction of the Berlin Wall in 1961; note that we look only at refugees who came to West Germany before 1951) being about 5 percent.

Table 5.2 provides sample means of the pre-war socio-demographic characteristics of local West Germans (column 1) and expellees (column 2) and the differences between the two groups (column 3).[11] We find very small, although significant differences between

[10] Retrospective data always bear the risk of misreported information. However, the German micro census does only ask major occupations which makes it quite likely that respondents remember correctly. Further, misreporting should not be correlated with expellee status which is the variable of interest in our analysis.

[11] We will provide more information about the GDR refugees at a later point in this chapter.

Table 5.2
Socio-Demographic Characteristics of Local West Germans and Expellees

	West Germans	Expellees	Difference
	(1)	(2)	(3)
Female	0.6038	0.5907	-0.0131***
	(0.489)	(0.492)	(0.003)
Age (1939)	31.6348	30.8424	-0.792***
	(9.784)	(9.821)	(0.069)
No Degree	0.0234	0.0302	0.0068***
	(0.151)	(0.171)	(0.001)
Elementary School	0.6425	0.6580	0.0155***
	(0.479)	(0.474)	(0.003)
Secondary School	0.2580	0.2391	-0.0189***
	(0.438)	(0.427)	(0.003)
High School	0.0126	0.0132	0.0006
	(0.112)	(0.114)	(0.001)
Technical School	0.0482	0.0446	-0.0036***
	(0.214)	(0.207)	(0.001)
University	0.0138	0.0132	-0.0006
	(0.117)	(0.114)	(0.001)
Real Estate (1939)	0.4847	0.5124	0.0278***
	(0.500)	(0.500)	(0.004)
Observations	146,786	23,183	

Notes: This table provides sample means of the pre-war socio-demographic characteristics of local West Germans (column 1) and expellees (column 2) and the differences between the two groups (column 3).

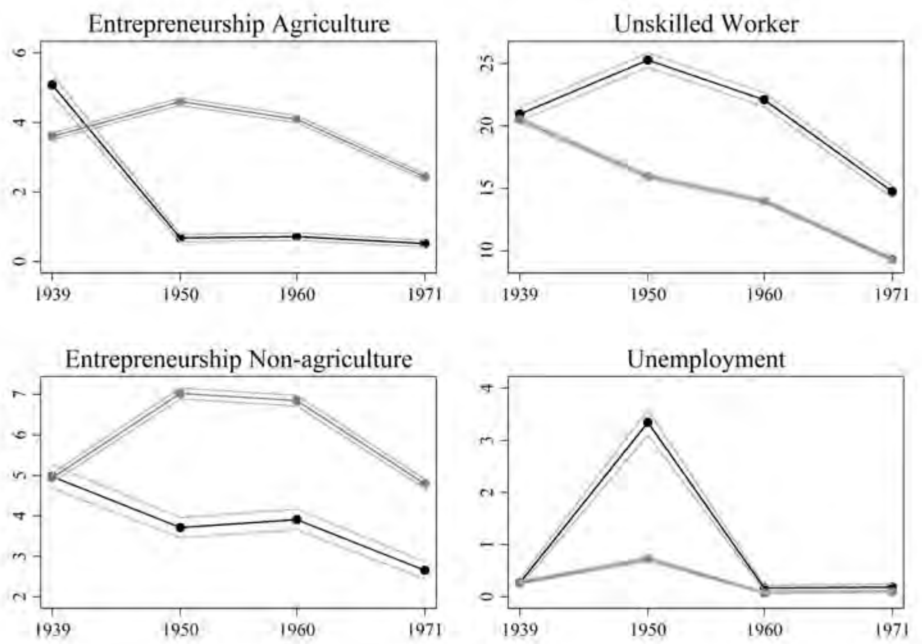
local West Germans and expellees in terms of demographic characteristics like gender, age, and educational attainment.[12] Virtually all local West Germans and expellees received at least basic schooling, about six percent in both groups completed advanced secondary education (high school or technical school), and more than one percent attended university. In 1939, a larger fraction of expellees owned property compared to local West Germans.

Figure 5.2 shows the occupational status of local West Germans and expellees before and after WWII (for more details see Table A5.1 in the Appendix). In 1939, the expellees' occupational structure is very similar to that of local West Germans. The most important difference is that expellees were more likely to work as self-employed farmers before WWII compared to local West Germans. This might also explain the larger fraction of expellees who owned real estate and worked in a family business as compared to West Germans

[12] We will control for the variables in all following regressions.

(see Table 5.2). We further observe that disproportionate numbers of expellees work in unskilled occupations after their displacement in 1950. In 1950, a smaller fraction of expellees reports to be self-employed or an entrepreneur (compared to local West Germans) and there are almost no self-employed farmers, presumably due to their loss of property.

Figure 5.2. Occupational Status of local West Germans and Expellees.



Notes: Graphs show the development (within 95 percent confidence bands) of four major occupations among expellees (black) and local West Germans (grey).

In accordance with the objectives documented in the Federal Expellee Law, we now focus on three distinct outcome categories. We construct dummy variables for each year indicating whether an individual (i) is unemployed or (ii) works in an unskilled occupation. Further, we consider if an individual is (iii) self-employed or an entrepreneur (i.e. owns a company with employees) where we distinguish between the agricultural and non-agricultural sector. We then separately estimate the following regression for all four years observed in our data to document the occupational development of the expellees compared to West Germans between 1939 and 1971.

$$Y_i = \alpha + \beta E_i + X'\gamma + \epsilon_i \tag{5.1}$$

Y_i represents one of the four outcome variables on the occupational status of individual i in a given year (1939, 1950, 1960, or 1971). E_i is a dummy variable indicating expellee status and the matrix X_i includes socio-demographic control variables such as gender, age, education and property ownership in 1939. The coefficient of interest is β which measures the difference in the outcome variable of interest between expellees and local West Germans conditional on the observable socio-demographic characteristics.

Since all outcome variables are binary, β can be interpreted as mean difference between expellees' and local West Germans' probability (in percentage points) of having a certain occupational status. Table 5.3 reports the coefficients from a linear probability model and marginal effects from a probit model on the expellee dummy for the four outcome variables in 1939 (before the displacement took place), in 1950 (after the displacement occurred, but before the Federal Expellee Law was introduced), and in 1960 and 1971 (after the Federal Expellee Law was introduced).[13] In 1939, there is no difference in the conditional probability to be unemployed between expellees and local West Germans (columns 1 and 2). After their displacement, expellees are about 2 percentage points more likely to be unemployed compared to local West Germans (columns 3 and 4). This difference decreases in 1960 (columns 5 and 6) and 1971 (columns 7 and 8). These figures clearly document that expellees had trouble to find jobs in West Germany right after their displacement. However, in 1960 and 1971 the difference in conditional probabilities of being unemployed between expellees and local West Germans decreased.

Row 2 in Table 5.3 shows the difference in the conditional probability to work in an unskilled occupation between expellees and West Germans in 1939, 1950, 1960 and 1971. In 1939, expellees are less likely to work in an occupation that does not require special training or education (columns 1 and 2). Similar to the expellees' increased conditional probability to be unemployed in 1950, expellees are about 8 percentage points more likely to be employed as an unskilled worker after their displacement than local West Germans (columns 3 and 4). One reason might be that expellees' school or job certificates were not accepted in West Germany or expellees were discriminated against by local firms or employers. Over time, the difference between expellees' and local West Germans' conditional probability to be employed as an unskilled worker decreased to 5.8 to 7.2 percentage points in 1960 (columns 5 and 6) and 2.7 to 4.5 percentage points in 1971 (columns 7 and 8).

[13] For the probit models, we report marginal effects evaluated at the means of the covariates.

Table 5.3

Difference in the Probability of Having a Certain Occupational Status between West Germans and Expellees

Occupation	1939		1950		1960		1971	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unemployed	-0.000	-0.000	0.026***	0.012***	0.001***	0.001***	0.001***	0.000***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unskilled worker	-0.004	-0.006**	0.085***	0.074***	0.072***	0.058***	0.045***	0.027***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
Entrepreneur (non-agricultural)	0.002	0.001	-0.033***	-0.034***	-0.030***	-0.031***	-0.023***	-0.021***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Entrepreneur (agricultural)	0.013***	0.006***	-0.043***	-0.033***	-0.038***	-0.025***	-0.022***	-0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. All regressions control for female, age, education and property in 1939. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Regarding self-employment and entrepreneurship in the non-agricultural sector there are no significant differences between expellees and West Germans in 1939. However, after their displacement, expellees are 3.3 and 3.4 percentage points less likely to be self-employed or entrepreneur in the non-agricultural sector than local West Germans (columns 3 and 4). Again, this difference gets absolutely smaller over time. The pattern for self-employed farmers looks slightly different. In 1939, expellees are 1.3 and 0.6 percentage points more likely to be self-employed in the agricultural sector (columns 1 and 2). After their displacement, the coefficient turns negative and again gets absolutely smaller over time.

Overall, Table 5.3 shows a consistent pattern. Except for self-employed farmers (and unskilled workers in the probit model), there are no significant differences in the occupational structure between expellees and local West Germans in 1939. Columns (3) and (4) reveal that the displacement presented a severe intervention for expellees; in 1950 expellees were more likely to be unemployed, more likely to work in an unskilled occupation, and less likely to be self-employed in both the non-agricultural and the agricultural sector than local West Germans. Relative to their pre-war situation, expellees suffered a loss in economic status. Columns (5) to (8) suggest for all outcome variables that expellees caught up with local West Germans. However, this pattern might only capture a general catch-up process of immigrants in a period of dramatic economic growth where unemployment among the local West Germans was close to zero instead of an effect of the law. We try to disentangle these two effects – catch-up and law-induced integration – by restricting our sample to expellees and local West Germans which are in the same occupational situation in 1950. We then infer the effects of the Federal Expellee Law by comparing their occupational status in 1960 and 1971.

The idea underlying our strategy is the following. Given the overall better economic situation of local West Germans in 1950, we assume that those West Germans who are in an occupational situation similar to the expellees are a selective group. We condition on some observable characteristics but there are many other factors that might explain their below-average situation. We consider the aggregate of these factors as an indication for a lower degree of integration which makes this subgroup of local West Germans more comparable to the expellees. Based on this argument, we use local West Germans in a similar occupational situation in 1950 as counterfactual group of individuals who are not integrated and not targeted by the law. Accordingly, a positive difference between these two groups should indicate a positive effect of the law. Given the selectivity of the subsample of local West Germans, we consider this approach as being pro-law as it should uncover even small effects.

5.4 Effects of the Federal Expellee Law

The economic development of expellees between 1939 and 1971 gives us an aggregate effect that captures the displacement effect, the effect of the law, and the general catch-up process of immigrants. The comparison of the coefficients in columns (1) and (3) or columns (2) and (4) of Table 5.3 documents the displacement effect on expellees' economic situation. To evaluate the effects of the law we look at the economic situation of expellees in 1960 and 1971, i.e. after the law was introduced. We try to disentangle the effects of the law from the catch-up effects by comparing expellees to specific subgroups of local West Germans. We then estimate the following regressions:

$$Y_{i,1960}|Y_{i,1950} = \alpha + \beta E_i + X'\gamma + \epsilon_i \tag{5.2}$$

$$Y_{i,1971}|Y_{i,1950} = \alpha + \beta E_i + X'\gamma + \epsilon_i \tag{5.3}$$

Y_i represents one of the four outcome variables on the occupational status of individual i either in 1960 or 1971. E_i is a dummy variable indicating expellee status, X_i are control variables namely gender, age, education and property ownership in 1939 and ϵ_i represents the error term that captures other unobservable effects. With respect to the objectives stated in the law, we condition expellees and local West Germans on their occupational situation ($Y_{i,1950}$) in 1950. When analyzing the restitution of previous or comparable occupations in case of degradation and the promotion of agricultural self-employment and entrepreneurship, we also condition on the occupational situation in 1939. By comparing expellees and local West Germans in very similar situations after WWII, we reduce a potential bias that may arise from unobservable characteristics.

5.4.1 Reduction of Unemployment among Expellees

To evaluate the effect of public employment services on the integration of expellees, we first compare the unemployment status of expellees and local West Germans who were unemployed in 1950 at two points in time after the introduction of the law (1960; 1971). Given that less than 1 percent of the local West Germans were unemployed in 1950, the comparison group of unemployed West Germans in 1950 is obviously selective. If anything, we should thus find a positive effect of the law when comparing unemployed expellees in 1950 to the selective group of unemployed local West Germans.

Table 5.4 reports the estimated coefficients from equations 5.2 and 5.3. We report the coefficients of linear probability models and marginal effects of probit models. The variable of interest is expellee status. If the law was successful in prioritizing unemployed expellees over unemployed local West Germans, we should find a negative and significant coefficient. However, the coefficient is close to zero and statistically not significant in both models. That is, compared to local West Germans expellees were not less likely to be unemployed in 1960 and 1971, given that both groups reported being unemployed in 1950. By taking into account the very low unemployment rates in both groups in 1960 and 1971, the reduction in unemployment among the expellees is likely driven by the overall growth of the economy and not by the Federal Expellee Law.

5.4.2 Restitution of Previous or Comparable Occupations

When analysing whether the law helped expellees find a job equivalent to their occupation prior to WWII, we compare expellees that reported that they were (i) not unemployed or working in an unskilled occupation in 1939 but (ii) worked in an unskilled occupation in 1950 (that is before the introduction of the law), with their local West German counterparts. We analyze whether expellees were more likely to work in a skilled job than their local West German counterparts after the introduction of the law, i.e. in 1960 and 1971. Again, the comparison group is highly selective since most local West Germans already found a job in 1950 which was equivalent to their occupation prior to WWII. If anything, it is thus most likely to find a positive effect of the law when comparing expellees to this selective group. If the law was successful in bringing expellees into their pre-war or equivalent occupations, we should find a negative, significant coefficient on the expellee dummy-variable. In other words, expellees should have a lower probability to work in an unskilled occupation compared to similar local West Germans. However, Table 5.5 shows that the coefficients are either very close to zero (columns 1 and 2) or even positive (columns 3 and 4). Beyond, both coefficients are statistically not different from zero. Even though Figure 5.2 documents that the fractions of expellees working in unskilled jobs decline over time, this is likely driven by general economic development and cannot be attributed to the law.

5.4.3 Promotion of Self-Employment and Entrepreneurship

A further goal of the law was to promote entrepreneurship and self-employment among expellees. The main mechanisms were (i) to reduce credit constraints for expellees who had to leave all their belongings in their homelands when coming to West Germany and (ii) to compensate them for a missing local social network that facilitates access to resources and customers. The latter was done by offering better opportunities to rent state-owned property for business purposes and ensuring that businesses run by expellees were treated preferentially when public contracts were awarded.

In agriculture, inheritance of the family farm is one of the key determinants of an individual's decision to become self-employed. The displacement eliminated this possibility for the expellees. Consequently, it would not be meaningful to compare expellees to local West Germans who are potential candidates for inheritance of a family farm. We thus restrict local West Germans and expellees to (i) individuals who worked in the agricultural sector in 1939, but not as family workers or self-employed farmers and (ii) individuals who were not self-employed farmers in 1950. These restrictions should ensure that we do not consider local West Germans who inherited a family farm between 1950 and 1960. We then analyze the effect of the law on agricultural self-employment in 1960 and 1971 by comparing those groups.

Table 5.6 documents that the law was not successful in promoting self-employment among expellees in the agricultural sector. Despite facilitated access to agricultural land, we find negative and significant coefficients on the expellee dummy in 1960 and 1971. In other words, compared to their West German counterparts who (i) had experience in working in agriculture, (ii) were not self-employed in agriculture in 1950 and (iii) were not likely to inherit a family farm, expellees were 1.7 to 2.4 percentage points less likely to

Table 5.4
Unemployment

	1960		1971	
	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)
Expellee	-0.002	-0.003	-0.004	-0.003
	(0.007)	(0.007)	(0.003)	(0.003)
Age	0.001*	0.001	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.044***	0.031***	-0.002	-0.001
	(0.011)	(0.006)	(0.004)	(0.003)
Elementary School	0.008	0.009	0.005**	0.035***
	(0.023)	(0.019)	(0.002)	(0.013)
Secondary School	0.014	0.015	0.007*	0.015***
	(0.023)	(0.019)	(0.004)	(0.006)
High School	-0.015		-0.000	
	(0.022)		(0.001)	
Technical School	-0.004	0.010	-0.000	
	(0.024)	(0.026)	(0.001)	
University	-0.016		0.002	
	(0.023)		(0.002)	
Property (1939)	-0.008	-0.008	-0.003	-0.003
	(0.007)	(0.007)	(0.003)	(0.002)
Unemployed in 1939	0.014	0.008	-0.006***	
	(0.029)	(0.015)	(0.002)	
Observations	1,837	1,778	1,837	1,631
(Pseudo-)R^2	0.020	0.073	0.003	0.053

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. The sample includes West Germans and expellees that reported to be unemployed in 1950. In the probit models, the dependent variable does not vary within some of the categories of the independent variables. There is, for example, only a very small number of observations with university degree and none of them are unemployed in 1960 or 1971. We drop those observations to fit the model. Constant is not reported. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 5.5
Unskilled Worker

	1960		1971	
	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)
Expellee	-0.001	-0.001	0.009	0.015
	(0.011)	(0.012)	(0.010)	(0.011)
Age	-0.009***	-0.009***	-0.021***	-0.026***
	(0.001)	(0.001)	(0.000)	(0.001)
Female	-0.180***	-0.189***	-0.196***	-0.214***
	(0.011)	(0.012)	(0.010)	(0.011)
Elementary School	-0.062**	-0.066**	-0.032	-0.039
	(0.028)	(0.032)	(0.025)	(0.028)
Secondary School	-0.191***	-0.203***	-0.121***	-0.133***
	(0.030)	(0.033)	(0.026)	(0.030)
High School	-0.431***	-0.444***	-0.351***	-0.570***
	(0.096)	(0.110)	(0.051)	(0.168)
Technical School	-0.321***	-0.334***	-0.206***	-0.234***
	(0.046)	(0.050)	(0.038)	(0.047)
University	-0.459*	-0.471	-0.141	-0.101
	(0.248)	(0.289)	(0.244)	(0.256)
Property (1939)	0.012	0.013	-0.004	-0.006
	(0.011)	(0.011)	(0.010)	(0.011)
Observations	8,439		8,439	
(Pseudo-)R^2	0.062	0.047	0.199	0.182

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. The sample includes West Germans and expellees that reported that they were not unemployed or working in an unskilled occupation in 1939 but worked in an unskilled occupation in 1950. Constant is not reported. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

become self-employed farmers in 1960 and 1971. Apparently, the law could not promote self-employment in the agricultural sector.

We next turn to self-employment and entrepreneurship in non-agricultural sectors. Since local West Germans might still have had a supportive local social network after WWII, we introduce an additional comparison group composed of GDR refugees who were not covered by the law. Like the expellees, GDR refugees had to leave their belongings and their local social network behind. However, Table 5.7 reveals that GDR refugees were 0.9 to 1.1 percentage points less likely to be self-employed and entrepreneurs in 1939 as compared to the expellees (columns 1 and 2). This lower entrepreneurial spirit in combination with their lack of assets and social networks qualifies them to be a suitable comparison group that, again, helps us to uncover even small effects of the law. In order to estimate the effect of the law on transition into entrepreneurship and self-employment, we compare entrepreneurial status in 1960 and 1971, respectively, (i) between expellees and West Germans who were not self-employed or entrepreneur in 1950 and (ii) between expellees and GDR refugees who were both not covered by the law and not self-employed or entrepreneur in 1950. Table 5.8 shows negative and significant coefficients on the transition into self-employment for expellees compared to both local West Germans and GDR refugees. That is, expellees were between 0.4 and 1.0 percentage points less likely to be self-employed in 1960 and 1971 given that they have not already been self-employed in 1950.

The comparison with local West Germans might be misleading, because it might be the case that the coefficients were even larger in absolute terms in absence of the Federal Expellee Law. However, the negative coefficients in the comparison of expellees and GDR refugees who (i) faced the same initial conditions in West Germany and (ii) seem to have a lower entrepreneurial spirit indicate that the promotion of self-employment and entrepreneurship among expellees was not successfully advanced by the law.

The law also aimed at reducing the risk of exit among self-employed expellees and entrepreneurs, for instance by transforming short-term loans into long-term contracts. Table 5.9 shows the probability to be self-employed or entrepreneur in 1960 and 1971, respectively, given an individual reported to be self-employed or entrepreneur in 1950. The negative coefficients in columns (1) to (4) show that expellees were substantially less likely to continue self-employed or entrepreneur in 1960 and 1971 as compared to their

Table 5.6
Self-Employed Farmers

	1960		1971	
	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)
Expellee	-0.022***	-0.017***	-0.024***	-0.017***
	(0.003)	(0.003)	(0.003)	(0.003)
Age	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.032***	-0.020***	-0.037***	-0.023***
	(0.003)	(0.003)	(0.004)	(0.003)
Elementary School	0.000	0.001	0.001	0.003
	(0.007)	(0.005)	(0.007)	(0.005)
Secondary School	-0.009	-0.005	-0.001	0.001
	(0.008)	(0.005)	(0.009)	(0.005)
High School	-0.023***		-0.024***	
	(0.008)		(0.009)	
Technical School	-0.001	0.001	0.019	0.007
	(0.016)	(0.006)	(0.019)	(0.006)
University	-0.035***		-0.038***	
	(0.008)		(0.008)	
Property (1939)	0.022***	0.013***	0.025***	0.013***
	(0.003)	(0.002)	(0.004)	(0.002)
Observations	6,730	6,691	6,730	6,691
(Pseudo-) R^2	0.027	0.152	0.033	0.169

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. The sample is restricted to local West Germans and expellees who worked in the agricultural sector in 1939, but not as family workers or self-employed farmers and who were not self-employed farmers in 1950. Some observations are dropped to fit the probit model. Constant is not reported. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 5.7
Difference in the Probability of Being Entrepreneur in the Non-agricultural Sector between the Three Groups

Group	1939		1950		1960		1971	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expellees to West Germans	0.002	0.001	-0.033***	-0.034***	-0.030***	-0.031***	-0.023***	-0.021***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Expellees to GDR refugees	0.011**	0.009**	-0.009*	-0.007**	-0.014**	-0.009***	-0.013***	-0.007***
	(0.005)	(0.004)	(0.005)	(0.003)	(0.006)	(0.003)	(0.005)	(0.002)
GDR refugees to West Germans	-0.011**	-0.010**	-0.025***	-0.021***	-0.018***	-0.014***	-0.015***	-0.009**
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)

Notes: This table reports coefficients on an expellee or east dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models, respectively. All regressions control for female, age, education, and property ownership in 1939. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 5.8
Entrepreneur Foundations

	West Germans				GDR refugees without expellee status			
Group	1960		1971		1960		1971	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expellee	-0.005***	-0.004***	-0.007***	-0.006***	-0.009**	-0.005***	-0.010**	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.004)	(0.001)
Age	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.012***	-0.009***	-0.015***	-0.010***	-0.012***	-0.008***	-0.014***	-0.009***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)	(0.002)	(0.001)
Elementary School	0.003**	0.004**	0.001	0.002	0.005**	0.009*	0.004*	0.006
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)	(0.002)	(0.005)
Secondary School	0.012***	0.011***	0.012***	0.010***	0.012***	0.014***	0.013***	0.012**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.003)	(0.005)
High School	0.020***	0.015***	0.026***	0.016***	0.020**	0.018***	0.029***	0.018***
	(0.004)	(0.003)	(0.005)	(0.002)	(0.009)	(0.006)	(0.010)	(0.006)
Technical School	0.034***	0.019***	0.034***	0.018***	0.044***	0.025***	0.040***	0.020***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.007)	(0.005)	(0.007)	(0.005)
University	0.030***	0.019***	0.045***	0.023***	0.035***	0.024***	0.046***	0.022***
	(0.005)	(0.003)	(0.006)	(0.002)	(0.012)	(0.006)	(0.013)	(0.006)
Property (1939)	0.005***	0.003***	0.006***	0.004***	0.004***	0.003***	0.005***	0.003***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)	(0.002)	(0.001)
Self-employed 1939	0.071***	0.025***	0.034***	0.017***	0.068***	0.020***	0.025***	0.011***
	(0.007)	(0.001)	(0.005)	(0.001)	(0.010)	(0.002)	(0.007)	(0.002)
Observations	158,797		158,789		24,112		24,112	
(Pseudo-)R^2	0.016	0.084	0.017	0.097	0.024	0.119	0.019	0.116

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. The sample includes West Germans and expellees, GDR refugees and expellees, respectively, that were not self-employed in 1950. Constant is not reported. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

Table 5.9
Entrepreneur Continuity

Group	West Germans				GDR refugees without expellee status			
	1960		1971		1960		1971	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expellee	-0.079***	-0.077***	-0.104***	-0.120***	-0.017	-0.016	-0.022	-0.025
	(0.015)	(0.014)	(0.016)	(0.018)	(0.045)	(0.049)	(0.051)	(0.056)
Age	-0.010***	-0.010***	-0.019***	-0.021***	-0.009***	-0.009***	-0.018***	-0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Female	-0.121***	-0.114***	-0.148***	-0.169***	-0.117***	-0.112***	-0.172***	-0.213***
	(0.010)	(0.009)	(0.010)	(0.012)	(0.036)	(0.034)	(0.034)	(0.044)
Elementary School	-0.085***	-0.085**	-0.039	-0.038	-0.018	-0.024	-0.314***	-0.392***
	(0.032)	(0.034)	(0.034)	(0.041)	(0.129)	(0.121)	(0.106)	(0.135)
Secondary School	-0.047	-0.048	0.032	0.042	0.137	0.124	-0.228**	-0.282**
	(0.032)	(0.034)	(0.034)	(0.041)	(0.128)	(0.121)	(0.107)	(0.135)
High School	-0.001	0.002	0.165***	0.186***	0.310**	0.334**	-0.094	-0.122
	(0.039)	(0.043)	(0.044)	(0.051)	(0.145)	(0.165)	(0.136)	(0.163)
Technical School	-0.022	-0.017	0.104***	0.118***	0.084	0.072	-0.231**	-0.287**
	(0.034)	(0.035)	(0.036)	(0.043)	(0.132)	(0.125)	(0.111)	(0.139)
University	0.060*	0.087**	0.316***	0.366***	0.283**	0.349**	0.105	0.089
	(0.035)	(0.039)	(0.038)	(0.046)	(0.131)	(0.141)	(0.116)	(0.148)
Property (1939)	0.032***	0.034***	0.038***	0.043***	0.056*	0.058*	0.130***	0.147***
	(0.008)	(0.008)	(0.009)	(0.010)	(0.031)	(0.031)	(0.030)	(0.035)
Self-employed 1939	0.084***	0.085***	0.031***	0.038***	0.073**	0.078**	0.020	0.026
	(0.009)	(0.009)	(0.010)	(0.011)	(0.030)	(0.032)	(0.031)	(0.037)
Observations	11,172		11,172		966		966	
(Pseudo-)R^2	0.066	0.0597	0.156	0.1234	0.092	0.0786	0.186	0.1562

Notes: This table reports coefficients on an expellee dummy for the linear probability models and marginal effects evaluated at the means of the covariates for the probit models. The sample includes West Germans and expellees, GDR refugees and expellees, respectively, that were self-employed in 1950. Constant is not reported. Robust standard errors are reported in parentheses. Significance level: * p<0.10, ** p<0.05, *** p<0.01.

West German counterparts. They were about 8 to 12 percentage points more likely to exit self-employment entrepreneurship compared to local West Germans.

Presumably, expellees that became self-employed or entrepreneur right after their arrival in West Germany did so out of necessity and as an alternative to being unemployed. Over time, they either were not successful and had to close their businesses or they did find more profitable employment. Since the nature of self-employment and entrepreneurship between West Germans and expellees who were self-employed in 1950 is likely to differ, the effect of the expellee law is difficult to assess by this comparison. We therefore make use of the GDR refugees that had the same starting conditions as expellees and compare them to expellees. If the law was successful in the support of self-employed or entrepreneurial expellees, we should find positive coefficients by comparing them to the group of refugees. However, we still find negative effects which are very close to zero and statistically not significant (cf. columns 5 to 8). That is, even compared to individuals who became self-employed or entrepreneur in a very similar situation as expellees, we do not find positive effects of the law on the continuation of entrepreneurship among expellees.

5.5 Conclusions

This paper evaluates the success of the 1953 Federal Expellee Law. It was designed to ameliorate the precarious situation of expellees upon their arrival in West Germany after WWII. The WWII shock was most severe for the group of expellees because they were forced to leave their homelands thus losing their possessions and their social ties. We allow for a general catch-up of expellees in a period of dramatic economic growth where unemployment among the local West German population was close to zero and compare the group of expellees to subgroups of local West Germans and GDR refugees who were arguably selective and possibly less likely to benefit from a general catch-up process during the economic boom in post-WWII Germany.

But despite the combination of economic boom and support by the law (which may upward-bias our estimations), we find no indication of a distinct effect of the law on expellees' situation. The possibility of expellees relocating to improve their economic conditions would just as well shift our estimates upwards. We therefore conclude that the improved economic situation of expellees can be attributed to the general economic boom in the aftermath of WWII and not to the provision of the Federal Expellee Law.

Table A5.1

Occupational Status of Local West Germans and Expellees

Occupation	Expellees				Local West Germans			
	1939	1950	1960	1971	1939	1950	1960	1971
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unemployed	0.27%	3.34%	0.17%	0.19%	0.27%	0.72%	0.08%	0.1%
Unskilled worker	20.94%	25.27%	22.09%	14.73%	20.47%	15.96%	13.99%	9.29%
Entrepreneur (agricultural)	5.08%	0.68%	0.72%	0.52%	3.6%	4.6%	4.09%	2.43%
Entrepreneur (non-agricultural)	4.97%	3.71%	3.91%	2.65%	4.94%	7.03%	6.85%	4.79%
Civil servant	3.21%	1.98%	2.2%	1.18%	2.67%	2.34%	2.05%	0.99%
Civil servant (upper- and upper-middle-level)	1.72%	1.35%	1.51%	0.91%	1.43%	1.38%	1.37%	0.75%
Pensioner, other non-employed	2.67%	9.71%	20.14%	7.95%	2.69%	5.73%	17%	8.09%
Housewife	26.85%	34.12%	29.87%	5.31%	29.4%	34.31%	30.14%	4.44%
Other	34.29%	19.84%	19.39%	66.56%	34.53%	27.93%	24.43%	69.12%

Notes: This table shows the percentage shares of expellees and local West Germans by occupational status before and after WWII. The category "Other" include employees, craftsmen, and family workers.

Chapter 6

Conclusions

The literature on the determinants of educational achievement has shown that the institutional set-up of schools and education systems plays a decisive role in the promotion of student learning. However, there is considerable uncertainty which institutional features increase overall educational performance. This thesis contributes to the literature on educational institutions and investigates the role of three aspects of education systems empirically. Conceptually, the analyses are based on the education production function that relates educational achievement indicators to the relevant determinants at the institutional level. Methodologically, microeconometric methods are employed to address the need for causal inference in policy-evaluations. In the following, the contribution to the literature will be pointed out and the key findings developed in this thesis will be summarized. To conclude, policy implications are derived.

Chapter 2 builds on the existing literature on school autonomy. While the overall evidence on the effects of school autonomy on student achievement is mixed, there are a few well-designed studies that report positive effects for European countries (see Barankay and Lockwood, 2007; Clark, 2009). In addition, several studies for developing countries actually find that positive effects of autonomy are restricted to non-poor areas (Galiani et al., 2008) or to school reform programs that simultaneously raised accountability (e.g., Gertler et al., 2012; Gunnarsson et al., 2009; Jimenez and Sawada, 1999). Based on this pattern, we investigate the heterogeneous effects of autonomy across levels of economic and educational development, and across different regimes of centralized accountability.

In contrast to other cross-country studies on these questions, our panel analysis allows us to control for time-invariant differences across countries by including country fixed effects.

Our central finding is that local autonomy has an important impact on student achievement, but this impact varies systematically across countries, depending on the level of economic and educational development. This implies that countries with otherwise strong institutions gain considerably from decentralized decision-making in their schools, while countries that lack such a strong existing structure may actually be hurt by decentralized decision-making. This result is most pronounced for autonomy in academic-content areas and for full school-level autonomy in contrast to joint decision-making between schools and external authorities. This result implies that the main effects of autonomy derive from independent decision-making at the school level. There is also evidence that autonomy works better in a system with generally well-functioning schools, measured by initial educational performance. Moreover, there is evidence that local decision-making complements the existence of external accountability systems that limit any opportunistic behavior of schools. This suggests that the level of accountability and the effectiveness of the education system are relevant channels through which the level of economic development affects the effectiveness of autonomy policies.

While we controlled for time-invariant differences across countries, the precise estimates that we report might be contaminated by other unobserved influences. Nevertheless, the qualitative pattern that we report is highly relevant and suggests that school autonomy does not make sense everywhere. As an overall result, this chapter has implications for the generalizability of findings across countries and education systems and emphasizes the importance of context-specific policies. The finding that school autonomy is heterogeneous across levels of development suggests that countries at different levels of economic and educational development face very different challenges in promoting educational achievement. Furthermore, the complementary of school autonomy to the existence of external accountability systems emphasizes the importance of the existing institutional setting when developing educational reforms.

Chapter 3 investigates the effect of single-sex schooling on student achievement. Despite a constant interest in whether single-sex schooling might have beneficial or adverse effects for boys and girls, the existing evidence is inconclusive. While many studies report positive effects, especially for girls, selection into single-sex schools is often not addressed in these studies. By exploiting the random assignment to schools in South Korea, we address the problem of selection and find that single-sex schooling has beneficial effects in math for girls, which are large in significance and magnitude. In contrast, there are neither beneficial nor adverse effects for boys. Subgroup analyses reveal that especially girls from low parental support backgrounds gain from single-sex schooling, with the benefits being larger for girls at the lower part of the performance distribution. This implies that those girls are – to some extent – harmed by the presence of boys in class and that highly supportive parents are able to compensate for these influences. The more general implication may be that girls with non-supporting parental backgrounds may be particularly influenced by less favorable peer characteristics in any school system.

Although arguments for and against single-sex education are well-developed, the underlying mechanisms are empirically not well documented. The exploration of potential channels at Korean middle schools shows that the positive effects for girls in math can neither be explained by differences in school and teacher characteristics nor by gender-tailored teaching styles and reduced gender stereotypes at single-sex schools. However, part of the effect can be attributed to a rougher classroom atmosphere at coeducational schools. This knowledge could be incorporated in the composition of classes and the allocation of school resources.

The fact that girls benefit from single-sex schooling but boys do not lose from the absence of girls at school suggests to organize learning in gender-segregated classes or schools. However, this result requires further investigation since other studies reveal that boys at coeducational schools benefit from larger shares of girls in class (see, e.g, Lavy and Schlosser, 2011). Moreover, the effect of single-sex schooling on a large number of other important outcomes, such as being capable of forming relationships with the opposite sex, is not well studied yet. Consequently, the separation of boys and girls in gender stereotypical subjects at otherwise coeducational schools could be a compromise. An overall implication of our results is that girls from disadvantaged families need particular attention, especially at coeducational schools.

Chapter 4 shows that program and course policies serve as incentives for students. While there is a growing literature on the effects of monetary incentives on academic achievement (see, e.g., Angrist et al., 2006; Fryer, 2011; Levitt et al., 2012), the question whether university program rules can be used as incentives has not received any attention yet. We find that the modification of program rules at a German university, which effectively doubled the time until students earn their certificates and which reduced the impact of each exam on the GPA, has a negative impact on students' test scores. While we cannot distinguish both effects, this result suggests that future research should focus on both dimensions, the timing of rewards and the leverage to students' effort choices. If further studies confirm their relevance, imminent milestones and a few powerful rather than many weak levers could be used as instruments to improve student achievement.

Furthermore, we find that an increase in the number of allowed resits increases the share of students that submitted blank papers. This indicates that students fail deliberately, so that they can resit the exam and improve. This finding suggests that on average students take more attempts to pass an exam if the number of allowed resits is not restricted. Involved with this policy is the fear that students could procrastinate and eventually drop out in their final semesters. Since our data does not allow to test this, this question requires further investigations.

Moreover, we show that students respond differently to the reforms depending on their ability. In particular, we find that poor students do not respond to changes in program rules, a finding which has already been documented in the literature on financial incentives and student achievement (Leuven et al., 2010). This result is consistent with the idea that especially low-performing students lack the ability to respond to incentives because they do not know how to translate their efforts into achievement (see also Levitt et al., 2012). An alternative implication would be that different groups of students have different objective functions, which need to be considered in designing incentive schemes. Overall, our research implies that policies are not just a necessary part of program implementation, but also offer universities a means of directing the effort that students put into their studies. Moreover, these policies are a promising field of research, because they are available to every university and, are, in contrast to monetary rewards, generally inexpensive.

Chapter 5 investigates the effect of an integration policy in the aftermath of World War II and contributes to the literature on the assimilation of migrants. Our results show that the 1953 Federal Expellee Law was not successful in improving the economic situation of expellees upon their arrival in West Germany. In particular, we find that reducing credit constraints and facilitating access to resources and customers did not promote the transition into entrepreneurship and self-employment of expellees neither in the agricultural nor in the non-agricultural sector. Similarly, we do not find positive effects of the law on the continuation of entrepreneurship among expellees. Furthermore, the decline of expellees working in unskilled jobs, and the reduction of the unemployment rate among expellees is likely to be driven by the general economic development and cannot be attributed to the law.

Our finding has two implications. First, many active labor market policies today put emphasis on public employment services to assist unemployed workers in finding jobs and entrepreneurship policies try to alleviate financing constraints of potential entrepreneurs by granting credits at reduced interest rates and tax credits. In this vein, our finding contributes to the literature on government employment and training programs that report at most modest benefits for program participants (see, e.g., Heckman et al., 1999). Second, the rise of armed conflicts, natural disasters, and large-scale infrastructure projects cause significant migration flows within countries. At the meantime, there is no economic research on policies to alleviate the consequences of these flows. Our results tentatively suggest that the integration of expellees through labor market policies is rather limited.

Overall, this thesis shows that educational institutions, in particular the locus of decision-making at schools (Chapter 2), the segregation of boys and girls at school (Chapter 3), and the organization of university programs (Chapter 4), play an important role in the promotion of student achievement. In addition, the success of a labor market policy aimed at the integration of expellees has been evaluated (Chapter 5). As an overall result, Chapter 2 has implications regarding the external validity of the within-country studies presented in Chapters 3, 4, and 5. While the analyses present important results and implications, the overall success of education policies also depends on a country's existing institutions and national peculiarities.

References

Ackermann, V. (1995). *Der "echte" Flüchtling. Deutsche Vertriebene und Flüchtlinge aus der DDR 1945 bis 1961*. Studien zur Historischen Migrationsforschung 1. Rasch-Verlag, Osnabrück.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., and Sapir, A. (2010). The Governance and Performance of Universities: Evidence from Europe and the US. *Economic Policy*, 25(61):7–59.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4):1167–1201.

Aktionsrat Bildung (2010). *Bildungsautonomie: Zwischen Regulierung und Eigenverantwortung*. VS Verlag für Sozialwissenschaften, Wiesbaden.

Alesina, A. F., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. T. (2003). Fractionalization. *Journal of Economic Growth*, 8(2):155–194.

Ammermüller, A. and Pischke, J.-S. (2009). Peer Effects in European Primary Schools: Evidence from PIRLS. *Journal of Labor Economics*, 27(3):315–348.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review*, 92(5):1535–1558.

Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review*, 96(3):847–862.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 1(1):136–163.

Angrist, J. and Lang, K. (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, 94(5):1613–1634.

Angrist, J. and Lavy, V. (2009). The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review*, 99(4):1384–1414.

Angrist, J. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, New Jersey.

Arcia, G., Macdonald, K., Patrinos, H. A., and Porta, E. (2011). School Autonomy and Accountability. System Assessment and Benchmarking for Education Results, World Bank, Washington, DC.

Barankay, I. and Lockwood, B. (2007). Decentralization and the Productive Efficiency of Government: Evidence from Swiss Cantons. *Journal of Public Economics*, 91(5-6):1197–1218.

Barrera-Osorio, F., Fasih, T., and Patrinos, H. A. (2009). Decentralized Decision-Making in Schools: The Theory and Evidence on School-Based Management. World Bank, Washington, DC.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics*, 106(2):407–443.

Barro, R. J. and Lee, J.-W. (1993). International Comparisons of Educational Attainment. *Journal of Monetary Economics*, 32(3):363–394.

Behrman, J. R., Choi, J., and Park, H. (2012). Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools. *Demography*, forthcoming.

Bethlehem, S. (1982). *Heimatvertreibung, DDR-Flucht and Gastarbeiterzuwanderung: Wanderungsströme und Wanderungspolitik in der Bundesrepublik Deutschland.* Klett Verlag, Stuttgart.

Beyer, S. and Bowden, E. M. (1997). Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. *Personality and Social Psychology Bulletin*, 23(2):157–172.

Bifulco, R. and Ladd, H. F. (2006). The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina. *Education Finance and Policy*, 1(1):50–90.

Bigler, R. S. and Signorella, M. L. (2011). Single-Sex Education: New Perspectives and Evidence on a Continuing Controversy. *Sex Roles*, 65(9-10):659–669.

Billger, S. M. (2009). On Reconstructing School Segregation: The Efficacy and Equity of Single-Sex Schooling. *Economics of Education Review*, 28(3):393–402.

Bishop, J. H. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *American Economic Review*, 87(2):260–264.

Bishop, J. H. (2006). Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems, and Peer Pressure. In Hanushek, E. A. and Welch, F., editors, *Handbook of the Economics of Education*, volume 2, pages 909–944. Elsevier.

Bishop, J. H. and Woessmann, L. (2004). Institutional Effects in a Simple Model of Educational Production. *Education Economics*, 12(1):17–38.

Blöchliger, H. and Vammalle, C. (2012). Reforming Fiscal Federalism and Local Government. OECD Fiscal Federalism Studies, OECD, Paris.

Booker, K., Gilpatric, S. M., Gronberg, T., and Jansen, D. (2007). The Impact of Charter School Attendance on Student Performance. *Journal of Public Economics*, 91(5-6):849–876.

Booth, A. L., Cardona, L., and Nolen, P. J. (2011). Gender Differences in Risk Aversion: Do Single-Sex Environments Affect Their Development? IZA Discussion Paper 6133.

Borjas, G. J. (1994). The Economics of Immigration. *Journal of Economic Literature*, 32(4):1667–1717.

Borjas, G. J. (1999). The Economic Analysis of Immigration. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3A, pages 1697–1760. Elsevier.

Brakman, S., Garretsen, H., and Schramm, M. (2004). The Strategic Bombing of German Cities during World War II and its Impact on City Growth. *Journal of Economic Geography*, 4(2):201–218.

Braun, S., Bauer, T. K., and Kvasnicka, M. (2012). The Economic Integration of Forced Migrants: Evidence for Post-War Germany. *Economic Journal*, forthcoming.

Brunello, G. and Rocco, L. (2011). The Effect of Immigration on the School Performance of Natives: Cross Country Evidence using PISA Test Scores. IZA Discussion Paper 5479.

Campbell, D. T. (1969). Reforms as Experiments. *American Psychologist*, 24(4):409–429.

Card, D. (1999). The Causal Effect of Education on Earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, pages 1801–1863. Elsevier.

Card, D. and Sullivan, D. (1988). Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment. *Econometrica*, 56(3):497–530.

Clark, D. (2009). The Performance and Competitive Effects of School Autonomy. *Journal of Political Economy*, 117(4):745–783.

Cochran, W. G. and Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhyā, Series A, Indian Journal of Statistics*, 35(4):417–446.

Cohen, A. (2012). Ew, Boys: The Brewing Legal Battle Over Same-Sex Education. TIME Ideas.

Coleman, J. S. (1961). *The Adolescent Society*. Free Press of Glencoe.

Coleman, J. S. (1966). *Equality of Educational Opportunity*. Government Printing Office, Washington, D.C.

Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics*, 24(4):761–786.

CREDO (2009). Multiple Choice: Charter School Performance in 16 States. Center for Research on Education Outcomes, Stanford University.

Dee, T. and West, M. (2011). The Non-Cognitive Returns to Class Size. *Educational Evaluation and Policy Analysis*, 33(1):23–46.

Durden, G. C. and Ellis, L. V. (1995). The Effects of Attendance on Student Learning in Principles of Economics. *American Economic Review*, 85(2):343–346.

Eisenkopf, G., Hessami, Z., Fischbacher, U., and Ursprung, H. (2011). Academic Performance and Single-Sex Schooling: Evidence from a Natural Experiment in Switzerland. CESIfo Working Paper 3592.

Epple, D. and Romano, R. E. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. In Jess Benhabib, A. B. and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1, pages 1053–1163. Elsevier.

Eurydice (2007). School Autonomy in Europe: Policies and Measures. Eurydice, Brussels.

Favara, M. (2011). The Cost of Acting “girly:” Gender Stereotypes and Educational Choices. University of Essex. Institute for Social and Economic Research.

Figlio, D. and Loeb, S. (2011). School Accountability. In E. A. Hanushek, S. M. and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, pages 383–423. Elsevier.

Frederick, S., Loewenstein, G., and O’Donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, 40(2):351–401.

Fryer, R. G. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *Quarterly Journal of Economics*, 126(4):1755–1798.

Fryer, R. G. and Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2):210–240.

Fuchs, T. and Woessmann, L. (2007). What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data. *Empirical Economics*, 32(2):433–462.

Galiani, S., Gertler, P., and Schargrodsky, E. (2008). School Decentralization: Helping the Good Get Better, but Leaving the Poor Behind. *Journal of Public Economics*, 92(10-11):2106–2120.

Galiani, S. and Perez-Truglia, R. (2011). School Management in Developing Countries. Paper presented at the conference education policy in developing countries: What do we know, and what should we do to understand what we don’t know? february 4-5. university of minnesota.

Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2012). College Cost and Time to Complete a Degree. Evidence from Tuition Discontinuities. *The Review of Economics and Statistics*, 94(3):699–711.

Gertler, P. J., Patrinos, H. A., and Rubio-Codina, M. (2012). Empowering Parents to Improve Education: Evidence from Rural Mexico. *Journal of Development Economics*, 99(1):68–79.

Gneezy, U. and Rustichini, A. (2000). Pay Enough or Don’t Pay at All. *Quarterly Journal of Economics*, 115(3):791–810.

Governor's Committee on Educational Excellence (2007). Students First: Renewing Hope for California's Future. Governor's Committee on Educational Excellence, Sacramento, CA.

Grosser, T. (2001). Die Aufnahme der Heimatvertriebenen in Württemberg-Baden und die regionalen Rahmenbedingungen ihrer Integration 1946 bis 1956. In Heumos, P., editor, *Heimat und Exil: Emigration und Rückwanderung, Vertreibung und Integration in der Geschichte der Tschechoslowakei*, pages 223–261.

Grosser, T. (2006). *Die Integration der Heimatvertriebenen in Württemberg-Baden (1945-1961).* Kohlhammer, Stuttgart.

Grützmacher, J., Ortenburger, A., and Heine, C. (2011). Studien- und Berufsperspektiven von Bachelorstudierenden in Deutschland. HIS: Forum Hochschule 7.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science*, 320(5880):1164–1165.

Gunnarsson, V., Orazem, P. F., Sánchez, M. A., and Verdisco, A. (2009). Does Local School Control Raise Student Outcomes? Evidence on the Roles of School Autonomy and Parental Participation. *Economic Development and Cultural Change*, 58(1):25–52.

Gustafsson, J.-E. (2006). Understanding Causal Influences on Educational Achievement through Analysis of Within-country Differences over Time. Invited Paper Presented at the 2nd IEA Research Conference, Washington. November 8-11.

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., Liben, L. S., and Martin, C. L. (2011). The Pseudoscience of Single-Sex Schooling. *Science*, 333 (6050):1706–1707.

Hanushek, E. A. (1970). The Production of Education, Teacher Quality, and Efficiency. In *Do Teachers Make a Difference?*, pages 79 – 99. U.S. Office of Education.

Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation using Micro Data. *American Economic Review*, 60(2):280–288.

Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Ressources*, 14(3):351–388.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(3):1141–1177.

Hanushek, E. A. (1994). Making Schools Work: Improving Performance and Controlling Costs. The Brookings Institution, Washington, DC.

Hanushek, E. A. (1996). Rationalizing School Spending: Efficiency, Externalities, and Equity, and Their Connection to Rising Costs. In Fuchs, V. E., editor, *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America*, pages 59–106. University of Chicago Press, Chicago, IL.

Hanushek, E. A. (2002). Publicly Provided Education. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4, pages 2045–2141. Elsevier.

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. *Economic Journal*, 113(485):F64–F98.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, 18(5):527–544.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., and Branch, G. F. (2007). Charter School Quality and Parental Decision Making with School Choice. *Journal of Public Economics*, 91(5-6):823–848.

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, Labor Force Quality, and the Growth of Nations. *American Economic Review*, 90(5):1184–1208.

Hanushek, E. A., Link, S., and Woessmann, L. (2012). Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA. *Journal of Development Economics*, forthcoming.

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

Hanushek, E. A. and Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3):607–668.

Hanushek, E. A. and Woessmann, L. (2011a). How Much Do Educational Outcomes Matter in OECD Countries? *Economic Policy*, 26(67):427–491.

Hanushek, E. A. and Woessmann, L. (2011b). The Economics of International Differences in Educational Achievement. In Eric A. Hanushek, S. M. and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, pages 89–200. Elsevier.

Hanushek, E. A. and Woessmann, L. (2012). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17(4):267–321.

Heckman, J., Lalonde, R., and Smith, J. (1999). The Economics and Econometrics of Active Labor Market Programs. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, pages 1865–2085. Elsevier.

Heidemeyer, H. (1994). *Flucht und Zuwanderung aus der SBZ/DDR 1945/1949 bis 1961: Die Fluchtlingspolitik der Bundesrepublik Deutschland bis zum Bau der Berliner Mauer.* Droste Verlag, Düsseldorf.

Hell, B., Weigand, S., Trapmann, S., and Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *German Journal of Educational Psychology*, 21(1):11–27.

Heyneman, S. P. and Loxley, W. A. (1983). The Effect of Primary School Quality on Academic Achievement Across Twenty-Nine High- and Low-Income Countries. *The American Journal of Sociology*, 88(6):1162–1194.

Hoffmann, D. (2000). Binnenwanderung und Arbeitsmarkt: Beschäftigungspolitik unter dem Eindruck der Bevölkerungsverschiebung in Deutschland nach 1945. In Hoffmann, D., Krauss, M., and Schwartz, M., editors, *Vertriebene in Deutschland: Interdisziplinäre Ergebnisse und Forschungsperspektiven*, pages 219–235, Munich. Oldenbourg.

Hoxby, C. M. (1994). Do Private Schools Provide Competition for Public Schools? National Bureau of Economic Research 4978, Cambridge, MA.

Hoxby, C. M. (1999). The Productivity of Schools and other Local Public Goods Producers. *Journal of Public Economics*, 74(1):1–30.

Hoxby, C. M. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. National Bureau of Economic Research 7867, Cambridge, MA.

Jackson, C. (2002). Can Single-Sex Classes in Co-Educational Schools Enhance the Learning Experiences of Girls and/or Boys? An Exploration of Pupils' Perceptions. *British Educational Research Journal*, 28(1):37–48.

Jackson, C. K. (2012). Single-Sex Schools, Student Achievement, and Course Selection: Evidence from Rule-based Student Assignments in Trinidad and Tobago. *Journal of Public Economics*, 96(5-6):173–187.

Jimenez, E. and Lockheed, M. E. (1989). Enhancing Girls' Learning Through Single-Sex Education: Evidence and a Policy Conundrum. *Educational Evaluation and Policy Analysis*, 11(2):117–142.

Jimenez, E. and Sawada, Y. (1999). Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program. *World Bank Economic Review*, 13(3):415–441.

Joshi, H., Leonard, D., and Sullivan, A. (2010). Single-Sex Schooling and Academic Attainment at School and Through the Lifecourse. *American Educational Research Journal*, 47(1):6–36.

Ju-Ho, L. (2004). The School Equalization Policy of Korea: Past Failures and Proposed Measure for Reform. *Korea Journal*, Spring:221–234.

Jürges, H. and Schneider, K. (2010). Central Exit Examinations Increase Performance...but take the Fun out of Mathematics. *Journal of Population Economics*, 23(2):497–517.

Kim, S. and Lee, J.-H. (2003). The Secondary School Equalization Policy in South Korea. KDI School Working Paper 02-05.

Kim, T., Lee, J.-H., and Lee, Y. (2008). Mixing versus Sorting in Schooling: Evidence from the Equalization Policy in South Korea. *Economics of Education Review*, 27(6):697–711.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to Learn. *Review of Economics and Statistics*, 91(3):437–456.

Lavy, V. and Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2):1–33.

Lavy, V., Silva, O., and Weinhardt, F. (2012). The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools. *Journal of Labor Economics*, 30(2):367–414.

Lee, V. E. and Bryk, A. S. (1986). Effects of Single-Sex Secondary Schools on Student Achievement and Attitudes. *Journal of Educational Psychology*, 78(5):381–395.

Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2010). The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

Leuven, E. and Sianesi, B. (2003). PSMATCH2: Stata Module to Perform full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing. Boston College Department of Economics, Statistical Software Components.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2012). The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. NBER Working Paper 18165, Cambridge, MA.

Loeb, S. and Strunk, K. (2007). Accountability and Local Control: Response to Incentives with and without Authority over Resource Generation and Allocation. *Education Finance and Policy*, 2(1):10–39.

Lüttinger, P. (1989). *Integration der Vertriebenen: eine Empirische Analyse*. Campus Verlag, Frankfurt/New York.

Madeira, R. (2007). The Effects of Decentralization on Schooling: Evidence from the Sao Paulo State's Education Reform. *Mimeo, University of São Paulo.*

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107(2):407–437.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

Marsh, H. W. (1989). Effects of Attending Single-Sex and Coeducational High Schools on Achievement, Attitudes, Behaviors, and Sex Differences. *Journal of Educational Psychology*, 81(1):70–85.

Metzler, J. and Woessmann, L. (2012). The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation. *Journal of Development Economics*, 99(2):486–496.

Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, 13(2):151–161.

Mourshed, M., Chijioke, C., and Barber, M. (2010). How the World's Most Improved School Systems Keep Getting Better. McKinsey and Company.

Müller, R. (1993). Der Staatsbeauftragte für das Flüchtlingswesen und die Anfänge der Flüchtlingsverwaltung in Württemberg-Baden. *Zeitschrift für Württembergische Landesgeschichte*, 52:353–399.

Mulligan, C. B. (1999). Galton versus the Human Capital Approach to Inheritance. *Journal of Political Economy*, 107(6):184–224.

MZU (1971). *Mikrozensus-Zusatzerhebung "Berufliche und soziale Umschichtung der Bevölkerung"*. GESIS.

Nechyba, T. J. (2003). Centralization, Fiscal Federalism and Private School Attendance. *International Economic Review*, 44(1):179–204.

Oates, W. E. (1972). Fiscal Federalism. Harcourt Brace Jovanovich, Inc., United States.

Oates, W. E. (1999). An Essay on Fiscal Federalism. *Journal of Economic Literature*, 37(3):1120–1149.

Organisation for Economic Co-operation and Development (1998). Reviews of National Policies in Education: Korea. OECD, Paris.

Organisation for Economic Co-operation and Development (2004). Education at a Glance: OECD Indicators 2004. OECD, Paris.

Organisation for Economic Co-operation and Development (2008a). Education at a Glace: OECD Indicators 2008. OECD, Paris.

Organisation for Economic Co-operation and Development (2008b). Women and Sustainable Development. OECD, Paris.

Organisation for Economic Co-operation and Development (2009). PISA 2009 Results: What Students Know and Can Do - Student Performance in Math, Science and Reading. OECD, Paris.

Organisation for Economic Co-operation and Development (2010). Education at a Glance: OECD Indicators 2010. OECD, Paris.

Ouchi, W. G. (2003). *Making Schools Work: A Revolutionary Plan to Get Your Children the Education They Need.* Simon & Schuster, New York, NY.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences. *Journal of Labor Economics*, 8(1):123–144.

Patrinos, H. A. (2011). School-Based Management. In Bruns, B., Filmer, D., and Patrinos, H. A., editors, *Making Schools Work: New Evidence on Accountability*, pages 87–140. World Bank, Washington, DC.

Pekkala Kerr, S. and Kerr, W. R. (2011). Economic Impacts of Immigration: A Survey. *Finnish Economics Papers*, 24(1):1–32.

Peterson, P. E., Howell, W. G., Wolf, P. J., and Campbell, D. E. (2003). School Vouchers: Results from Randomized Experiments. In Hoxby, C. M., editor, *The Economics of School Choice*, pages 107–144. University of Chicago Press, Chicago.

Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development*, 22(9):1325–1343.

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2):111–134.

Riordan, C. (1990). *Girls and Boys in School: Together or Separate?* Teachers College Pres, New York, NY.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence From Panel Data. *American Economic Review Papers and Proceedings*, 94(2):247–252.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2):681–704.

Sarvimäki, M., Uusitalo, R., and Jäntti, M. (2009). Long-Term Effects of Forced Migration. IZA Discussion Paper 4003.

Schlotter, M., Schwerdt, G., and Woessmann, L. (2011). Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide. *Education Economics*, 19(2):109–137.

Schmidt, C. M. (1994). The Economic Performance of Germany's East European Immigrants. CEPR Working Paper 963.

Schneeweis, N. (2010). Educational Institutions and the Integration of Migrants. *Journal of Population Economics*, 24(4):1281–1308.

Schraut, S. (1995). Flüchtlingsaufnahme in Württemberg-Baden 1945 bis 1949. In *Amerikanische Besatzungsziele und demokratischer Wiederaufbau im Konflikt*, Munich. Oldenburg.

Schütz, G., Ursprung, H. W., and Woessmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos*, 61(2):279–308.

Smith, J. and Todd, P. (2005). Does Matching Overcome LaLonde's Critique of Non-Experimental Estimators? *Journal of Econometrics*, 125(1-2):305–353.

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Thompson, J. S. (2003). The Effect of Single-Sex Secondary Schooling on Women's Choice of College Major. *Sociological Perspectives*, 46(2):257–278.

Todd, P. E. and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function For Cognitive Achievement. *Economic Journal*, 113(485):F3–F33.

Vaskovics, L. A. (2002). *Gesellschaftliche Desorganisation und Familienschicksale*. Iudicum-Verlag, Munich.

Weiß, M. (2004). Wettbewerb, Dezentralisierung und Standards im Bildungswesen. In Federal Ministry for Education and Research, editor, *Investitionsgut Bildung*. Federal Ministry for Education and Research, Berlin.

West, M. R. and Woessmann, L. (2010). Every Catholic Child in a Catholic School: Historical Resistance to State Schooling, Contemporary Private Competition, and Student Achievement Across Countries. *Economic Journal*, 120(546):F229–F255.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review Papers and Proceedings*, 95(2):199–203.

Woessmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: The International Evidence. *Oxford Bulletin of Economics & Statistics*, 65(2):117–170.

Woessmann, L. (2005). The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, 13(2):143–169.

Woessmann, L. (2007). International Evidence on Expenditure and Class Size: A Review. *Brookings Papers on Education Policy*, 2006/2007:245–272.

Woessmann, L. (2008). How Equal are Educational Opportunities? Family Background and Student Achievement in Europe and the United States. *Zeitschrift für Betriebswirtschaft*, 78(1):45–70.

Woessmann, L., Luedemann, E., Schuetz, G., and West, M. R. (2009). *School Accountability, Autonomy, and Choice around the World.*, volume 4. Elgar, Cheltenham.

World Bank (2004). World Development Report 2004: Making Services Work for Poor People. World Bank, Washington, DC.

0.92

Curriculum Vitae

Susanne Link
born 25.12.1984 in Mosbach

Since 08/2010	Junior Economist at the Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
09/2009 – 05/2013	Ph.D. in Economics Ludwig-Maximilians-Universität, Munich (Degree: Dr. oec. publ.)
04/2005 – 07/2009	Studies of Economics Ludwig-Maximilians-Universität, Munich (Degree: Diplom-Volkswirt)
07/2007 – 12/2007	Visiting Student University of Sydney, Australia
07/2004	Abitur Auguste-Pattberg-Gymnasium, Mosbach