Empirical Studies on Aspects of Education Inequality in Germany

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation

von Diplom-Volkswirtin, Master of Arts Andrea Maria Mühlenweg

geboren am 27.03.1977 in Karlsruhe

2007

Erstgutacher: Prof. Dr. Patrick Puhani

Zweitgutachter: Prof. Dr. Horst Entorf

Tag der Disputation: 08.11.2007

Diese Arbeit widme ich Hans-Jörg,
meinen Eltern und der ganzen Familie

Acknowledgment

First and foremost, I would like to thank my supervisor Patrick Puhani for his continuous guidance, encouragement and support during the development of this thesis. His expertise, insights and suggestions shaped my research skills and prepared me for future challenges. Throughout my doctoral work he encouraged me to develop independent project ideas. I am grateful to him and to my co-advisor Horst Entorf for helpful advices and inspiring discussions during these last three years.

I am indebted to colleagues at Darmstadt University of Technology and the Leibniz University of Hannover for valuable discussions. The different chapters of this thesis benefited from valuable comments and support of various persons and institutions who and which are mentioned and thanked at the beginning of the respective studies. Especially Christian Dustmann, Bernd Fitzenberger, Gianno De Fraja, Peter Fredriksson, Knut Gerlach, Szilvia Hamori, Olaf Hübler, Michael Lechner, Edwin Leuwen, Stephen Machin, Sandra McNally, Uwe Jirjahn, Christian Pfeifer, Sergiy Radyakin, Regina Riphahn, Kjell Salvanes, Claus Schnabel, Reinhold Schnabel, Kerstin Schneider, Uta Schönberg and participants in the Ph.D. seminars in Applied Labour Economics at the University of Hannover, Darmstadt University of Technology and the ZEW-Rhein-Main-Neckar-Seminar provided valuable comments to me or the co-author of the two included joint papers at conference or workshop presentations. Thanks to Philip Savage and Björn Schumacher for excellent research assistance.

Financial support from the Anglo-German Foundation within the project 'The Economics and Politics of Employment, Migration and Social Justice' of the initiative 'Creating Sustainable Growth in Europe' is gratefully acknowledged for parts of the projects. Thanks to Hans-Peter Hafner of the Research Data Center (Forschungsdatenzentrum) of the Statistical Office of the state of Hessen for his cooperation and help with the administrative data. I am obliged to the school principal, teachers, pupils and parents as well as persons at the local school authority for their support of the school intervention study.

I am especially grateful to my colleagues and friends Anabell and Oliver for sharing a lot of cups of coffees and cookies and thus increasing my labour productivity. Similarly, I extend many thanks to my sister Christiane and her husband Ulli as well as to my friends Ute, Bernadette, Claudia, Mareike and Lana. I am grateful to Jochen, who, about four years ago, pointed out to me that people at Darmstadt University of Technology might be interested in empirical education research and thus made me being interested in this place.

Finally, but most importantly, I would like to thank my husband Hans-Jörg for being so understanding and patient during these years and especially the last three weeks. To him and my parents I owe a debt of gratitude for their support and motivation and for being a wonderful family.

Abstract

This dissertation considers aspects of education inequality in Germany. The German education system is known to produce relatively high education inequality, particularly as a result of selecting pupils into secondary school tracks when they are about ten years old. The studies presented discuss different sources with the potential to increase flexibility and to decrease inequality in education. Firstly, the establishment of so-called ‘support stages’, which delay the timing of tracking for two years (to seventh grade), is examined. Furthermore, German school entry-age regulations are considered, where flexibilities related to the track choice may again reduce initial disadvantages of early school entrants. Finally, a school intervention project examines the benefits of single-gender education. This project aims to investigate whether gender specific preferences related to technical subjects may be changed through such an intervention.

Keywords: education, inequality, identification

Kurzzusammenfassung

Die vorliegende Dissertation untersucht verschiedene Aspekte der Bildungsungleichheit in Deutschland. Das deutsche Bildungswesen mit dem selektiven, dreigliedrigen Sekundarschulsystem hat den Ruf, eine hohe Bildungsungleichheit zu produzieren. Diese Arbeit weist verschiedene Flexibilitätspotentiale auf, die abschwächend auf die starke Selektivität des Systems und die beobachtete Bildungsungleichheit wirken könnten. So wird zunächst die Wirkung der Förderstufen untersucht, welche die Zuweisung auf die unterschiedlichen Sekundarschulformen um zwei Jahre (auf die siebte Klasse) verschieben. Außerdem wird die Auswirkung gegenwärtiger Einschulungsregelungen beleuchtet, wobei wiederum Flexibilitäten im Schulsystem frühe Nachteile, die jung eingeschulten Kindern entstehen, ausgleichen können. Ferner wird die Möglichkeit des zeitweise getrennt geschlechtlichen Unterrichts diskutiert; dabei wird im Rahmen eines Schulprojektes der Frage nachgegangen, ob fächerspezifische Unterschiede von Jungen und Mädchen durch eine solche Unterrichtsorganisation beeinflusst werden können.

Schlagworte: Bildung, Ungleichheit, Identifikation

Contents

List of Tables

List of Figures

List of Abbreviations

Aug	August
BW	German state of Baden-Württemberg
CEPR-IFAU	Centre for Economic Policy Research – Institute for Labour Market Policy Evaluation
cf.	Confer
ed.	Editor
eds.	Editors
e.g.	Exempli gratia
et al.	Et alii
etc.	Et cetera
EUR	Euro
Exo	Vector of exogenous variables
F	F-statistic
GSOEP	German Socio-Economic Panel Study
i.e.	Id est
IEA	International Association for the Evaluation of Educational Achievement
IZA	Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor)
Jul	July
KW	Kruskal-Wallis test
LATE	Local Average Treatment Effect
N	Numbers (of observation)
n.a.	Not available
obs. / observ.	Observations
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
p.	Page
Ph.D.	Doctor of Philosophy
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PISA-E	National extension study to PISA
s.d.	Standard deviation
s.e.	Standard error
Sep	September
TIMSS	Third International Mathematics and Science Study
U.K.	United Kingdom
U.S.	United States

U.S.A.	United States of America
vs.	Versus
2SLS	Two-Stage Least Squares
ZA	Zentralarchiv für Empirische Sozialforschung (Central Archive for Empirical Social Research)
ZEW	Zentrum für Europäische Wirtschaftsforschung (Centre for European Economic Research)

Introduction

"I have no data yet. It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

Sherlock Holmes, in 'A Scandal in Bohemia'

The various chapters of this dissertation are the result of separate empirical research projects shedding light on aspects of education inequality in Germany. All papers rely heavily on available data. Limitations of the insights gained are mainly due to issues of data availability. Thus, empirical research proceedings bear a marked resemblance to Sherlock Holmes's work, where the search for useful data is crucial. In fact, one innovation of this dissertation is that it draws on newly available administrative data covering all students in one German state (cf. Chapters 1-3). Furthermore, Chapter 4 relies on the strategy of conducting a new school intervention project in order to collect the required data.

The German education system is known to produce relatively high education inequality, particularly as a result of selecting pupils into secondary school tracks when they are about ten years old. All the studies presented discuss different sources with the potential to increase institutional flexibility and to decrease inequality in education. The establishment of so-called 'support stages' (*Förderstufe*) in the German state of *Hessen,* for example, delays the tracking decision for two more years. Chapter 1 compares the traditional German early-tracking system to these later tracking institutions. Empirical findings from this study suggest that later tracking is favourable for pupils with a disadvantaged social background and produces less education inequality. However, there is also some evidence of negative effects of later tracking for pupils on top of the (conditional) performance distribution.

Aspects of the German tracking system are also the topic of Chapter 2 and Chapter 3. These two studies consider the impact of pupils' age at school entry on their later educational performance. The studies ask the question whether pupils who are born earlier in the year and thus enter school relatively young due to the official school entry age regulation in Germany are disadvantaged in the education system. Chapter 2 demonstrates that, in fact, pupils who are about a year older than their peers when they enter school perform significantly better in the fourth grade of elementary school and have a higher probability of being tracked to the highest level secondary school (*Gymnasium*). Chapter 3 examines whether these school entry-age effects persist up to the senior high school years. Again, this study is based on data for the German state of *Hessen*, which is known to allow for a relatively high mobility between school tracks. In fact, the empirical analysis reveals that the school entry age effect is neutralized by the possibility of track upgrading after tenth grade.

Chapter 4 focuses on gender differences in education. German school laws allow for the possibility to instruct male and female pupils in separate classes. A school intervention project aims at answering the question whether gender specific preferences related to technical subjects may be changed through single-gender education. Findings from this project are not conclusive. The project fails to identify positive effects from single-gender education but the interpretation is impeded by several confounding factors.

All in all, the different studies suggest that flexibilities in the education system allowing deviation from traditional principles of education in Germany (such as early tracking or school entry-age regulations) provide important channels for reducing education inequality. In addition, scope for further research is pointed out, particularly related to the impacts of single gender education.

Chapter 1

Educational Effects of Alternative Secondary School Tracking Regimes in Germany *

* This chapter is based on earlier discussion paper versions: Darmstadt Discussion Paper in Economics, 176, 2006 and Discussion Paper of the Department of Economics, Leibniz Universität Hannover, 353, 2006.

Abstract: This chapter examines educational outcomes of pupils selected to secondary school types by different tracking regimes in a German state: Pupils are alternatively streamed after fourth grade or after sixth grade. Regression results mainly indicate that, estimated on the mean, there are no negative effects of later tracking on educational outcomes in the middle of secondary school. Notably, positive effects are observed for pupils with a less favorable family background. Quantile regressions reveal that the effects of later tracking are positive for the lower quantiles but decrease monotonically over the conditional distribution of test scores, turning into significant negative effects for the upper quantiles. Thus, the findings suggest that later tracking reduces education inequality, but to the detriment of pupils with an advantaged educational background.

JEL classification: I21, I28

Keywords: education, segregation, immigration

Acknowledgement: Financial support from the Anglo-German Foundation within the project 'The Economics and Politics of Employment, Migration and Social Justice' of the initiative 'Creating Sustainable Growth in Europe' for parts of the study is gratefully acknowledged. I am grateful to Horst Entorf, Knut Gerlach, Szilvia Hamori, Uwe Jirjahn, Christian Pfeifer, Patrick Puhani, Sergiy Radyakin, Philip Savage and three anonymous referees as well as participants of the Ph.D. Seminars in Applied Labor Economics at the University of Hannover and the Darmstadt University of Technology for valuable comments. I also thank Hans-Peter Hafner from the Research Data Center (Forschungsdatenzentrum) of the Statistical Office of the state of *Hessen* for help with the administrative data for *Hessen*. Thanks are due to Ursula Dörger from the Ministry of Education in *Hessen* who kindly answered my questions concerning school regulations. All remaining errors in this study are my own.

Table of Contents for Chapter 1

1.1 Introduction to Chapter 1

Numerous European countries select pupils into more or less academic tracks at some point during their secondary education. The rationale behind educational tracking or streaming is to provide a homogeneous learning environment which is supposed to foster specific pupils' abilities and to improve educational outcomes. From a theoretical point of view, the educational setup with respect to tracking may be considered as the result of an optimization process. Thus, recent studies by Brunello *et al.* (2007) and Ariga *et al.* (2005) model optimal tracking time as being determined by a trade-off between negative and positive effects of early tracking: The negative effect stems from the assumption that the tracking decision is the more appropriate (with respect to actual, unobserved individual ability) the later tracking takes place. The counteracting positive effect is due to more able pupils benefiting from a more selective system.[1]

In Germany, pupils are generally tracked into three different types of secondary schools at a relatively early point of their educational careers (mostly at the age of ten). Track choice mainly depends on the decisions made by parents. Recently, researchers have argued that this early tracking regime is an important source of high educational inequality: For example, Dustmann (2004) states that early tracking enforces intergenerational immobility because of strong influences of parental views on the children's (early) educational decision. The study shows that parental education and occupational status have a significant impact on the children's secondary school choice and subsequent educational attainment in Germany. In addition, these parental influences yield to differences in the children's earnings later in life. These views are confirmed by recent studies mainly drawing on internationally standardized test score data for different countries: The cross-county comparisons by Hanushek and Wößmann (2006), Entorf and Lauk (2006), Ammermüller (2005), and Schütz *et al.* (2005)[2] and the Swiss cross-canton study by Bauer and Riphahn (2006) indicate that countries featuring tracking and especially early tracking systems are characterized by relatively high educational inequality and lower average performance. Pekkarinen (2005) shows that later tracking

[1] Non-linear peer-effects are assumed in these models. Epple *et al.* (2002) is a further study modelling implications of school tracking. However, this paper refers to the somewhat different context of ability tracking within public and private schools. Different selection mechanisms to school tracks are examined in Fernandez (1998).

[2] The empirical paper by Schütz *et al.* (2005) also offers a theoretical model linking the timing of tracking to education inequality.

yields higher gender differences in education in favor of girls and decreases the subsequent gender wage gap.[3]

One special feature of the German educational system is that besides the traditional early tracking schools some later tracking schools exist, too: In so-called 'support stages' (*Förderstufe*) or 'orientation stages' (*Orientierungsstufe*) tracking is postponed for two years. The idea is that pupils are given more time to develop specific skills and interests and that teachers and parents receive improved information for the transition decisions to secondary schools. To date and to my knowledge, no empirical research has been undertaken to identify a causal effect of the 'support stages' on educational outcomes using appropriate statistical strategies.[4]

This study aims at examining educational effects of these special schools in one German state (*Hessen*), for which data on the entire pupil population is available. The central methodological problem when comparing educational outcomes by tracking regime is that tracking regime choice is endogenous to educational outcomes. Thus, estimates of the 'support stage effect' are likely to be biased in a simple regression framework. In brief, since the endogeneity bias can be considered to be an omitted variable bias, I examine how the estimated effect changes whilst a broad variety of background characteristics is controlled for.

Since some of the above mentioned studies demonstrate (based on comparisons of different countries) that later tracking reduces educational inequality, this paper also focuses on inequality aspects of the tracking regime. First of all, regression results are presented for different sub-groups according to pupils' family background. Furthermore, quantile regressions demonstrate the difference of the later tracking effect for pupils at different quantiles of the conditional performance distribution.

One additional feature of my research is that I use newly available administrative data on the entire pupil population for one German state (*Hessen*). To my knowledge, this data base covering four waves of data has not been used before in empirical research studies (with the exception of Puhani and Weber, 2007a and Puhani and Weber, 2007b). Therefore, this is the first study providing detailed information on the importance of alternative tracking types in a German state based on individual level data.

---

[3] While the focus of the present paper is on tracking of pupils to academic and vocational school types further empirical studies consider ability grouping within schools. Recent papers examining this version of tracking are, for example, Zimmer (2003), Figlio and Page (2002) and Betts and Shkolnik (2000).

[4] An early study of the 'support stages' in *Hessen* is provided by Hopf (1979) and describes the development and organisation of the schools as well as experiences of parents, teachers and pupils in this school type. The study does not compare 'support stage' outcomes to outcomes of alternative school types using evaluation techniques. A similar approach is taken in the studies of 'orientation stages' in *Bremen* by Eiko (1989) and Eiko (1991). Henze *et al.* (1996) focuses on low ability pupils within 'orientation stages' in the state of *Niedersachsen*.

This chapter is organized as follows: Section 1.2 describes the German education system with an emphasis on the institutional framework of the state of *Hessen*. Section 1.3 provides descriptive evidence on tracking in *Hessen*. It is shown that pupils having attended later tracking schools perform worse (in terms of the secondary education level reached) than pupils who have been tracked early. However, these results are driven by the endogeneity of regime choice. The methodological framework for an analysis of track choice taking its endogeneity with respect to educational outcomes into account is introduced in Section 1.4 together with the results: Overall, there seem to be no negative effects of later tracking. However, sub-group analyses and quantile regression results reveal that 'support stages' seem to work in favor for children with a disadvantaged education background whilst there are negative effects on pupils on top of the conditional performance distribution. Thus, later tracking may, in fact, decrease education inequality but to the detriment of the top performers. Section 1.5 discusses the findings and presents conclusions.

1.2 Stylized Facts

1.2.1 Institutional Background

Traditionally, the German school system is characterised by early ability streaming of pupils. Table 1.1 provides an overview of the tracking systems in selected industrialised countries:[5] While many European countries track pupils to more or less academic secondary school types, Germany's regular tracking age of ten is rather early in international comparison. To be more specific, in Germany pupils are selected into three school types after four years of elementary school:[6] The most 'able' pupils are supposed to attend the *Gymnasium*, which is a nine- (or eight-) year higher-level secondary school and enables pupils to pursue further academic studies (*e.g.* at universities).[7] An alternative school track is offered by the *Realschule* as an intermediate level secondary school which generally lasts six years and prepares pupils for a rather vocational education. Finally, the *Hauptschule*, as the lowest level secondary school type, is supposed to be the most vocational and least academic track and lasts five years. In principle, it is possible to change tracks after the initial track decision. However, different

[5] Besides explicitly streaming pupils to vocational and academic tracks, in some countries it is common to select pupils to different classes within comprehensive secondary schools according to ability (as it is the case in the U.S.). This version of tracking is not considered in Table 1.1.
[6] In the East German states Berlin and *Brandenburg*, primary school generally covers six grades.
[7] Recently there has been a tendency to shorten the duration to eight years. In the East German states *Sachsen* and *Thüringen*, the higher secondary school generally takes eight years.

curricula for the different school types complicate switching tracks, especially after sixth grade.[8]

Besides the system of streaming pupils to the different secondary school types after fourth grade, later tracking school types also exist. These school types, which are called 'support stages' (*Förderstufe*) or 'orientation stages' (*Orientierungsstufe*), track pupils after sixth grade. Later tracking schools were mainly introduced in different regions at the end of the 1950s and in the 1970s:[9] Especially in the 1950s, educational experts developed the idea of so-called 'support stages'.[10] While the traditional elementary schools were to be maintained, the *Förderstufe* sought to combine grades five and six in an autonomous comprehensive school type which would be located at traditional German lower secondary or primary schools. In the states of *Hessen* and *Niedersachsen*, this school type was introduced on a larger scale alongside the traditional tracking system.[11] Reasons for introducing 'support stages' may have been rather theoretical ones (*e.g.* to foster equal educational opportunities) or practical ones: Schools in rural areas tended to introduce 'support stages' so that all fifth and sixth graders could be provided with local secondary education.[12]

All in all, discussions on the idea of prolonged comprehensive schooling generated a mixed system of institutions in Germany: The state of *Hessen* introduced the offer of 'support stages' (*Förderstufe*) in some schools coexisting with the traditional selective school types. Children in these 'support stage' schools are normally taught in comprehensive classes, while separate classes according to ability may exist for mathematics and the first foreign language (mostly English).

Concerning the regulations in the other German states, in most states, pupils are mainly still selected to different secondary school types after fourth grade. Furthermore, the states of *Bremen* and *Niedersachsen* used to have fully established comprehensive 'orientation

[8] Hardly any figures on switching tracks exist. Baumert *et al.* (2003) states that 14.4 % of German 15- year-old pupils in the PISA study claim to have switched from initial secondary school track to another track. Pischke (2003) explains that 7 % of pupils switched to higher level schools from lower or intermediate secondary schools in 1966. Recent evidence based on administrative data for *Hessen* is given in Puhani and Weber (2007b) demonstrating that track upgrades are more frequent than downgrades.

[9] For further information on the history of comprehensive secondary schooling see Hessisches Kultusministerium (1995) and Jürgens (1991).

[10] This idea was developed in the '*Rahmenplan zur Umgestaltung und Vereinheitlichung des allgemeinbildenden öffentlichen Schulwesens*' of the *Deutscher Ausschuß für das Erziehungs- und Bildungswesen* in 1959.

[11] The first *Förderstufe*-type school was already introduced in 1955 in *Hessen* in the so-called *Schuldorf Bergstraße*. Whether a 'support stage' was introduced at a specific school was instigated by the school authority (*Schulträger*) and the respective school.

[12] A further discussion of the idea of prolonged comprehensive schooling emerged after the formation of the 'German Education Council' (*Deutscher Bildungsrat*) in 1965. In 1970, the council suggested that a comprehensive 'orientation stage' following the four years of elementary school should cover grades five and six. This is especially documented in the '*Strukturplan für das Bildungswesen*' from 1970. In the following years, representatives of all German *Länder* in the *Bund-Länder-Kommission* discussed how to organise this new school type. However, the projected system of homogenous nation-wide 'orientation stages' could not be enforced.

stages' covering grades five and six but abolished them in 2005 and 2004 respectively. It is only in *Berlin* and *Brandenburg* that elementary school traditionally takes six instead of four years.

In addition, general comprehensive schools exist in Germany, too. Pupils in the former German Democratic Republic used to be taught in comprehensive schools (*Einheitsschule*) until tenth grade. In West Germany, comprehensive schools (*Gesamtschule*) were introduced as an 'experiment' in several schools in the 1960s and lead to grade 10 or 13 respectively. From 1973 to 1982 all German states introduced some experimental comprehensive schools. Pupils in comprehensive schools are taught in different ability groups (only) in some subjects (*integrierte Gesamtschule*) or are allocated to an internal track according to their ability similar to the traditional school tracks (*kooperative Gesamtschule*). Nowadays, the acceptance of comprehensive schools largely varies between the German states: While there is only one comprehensive school left in Bavaria (as a remnant of the nation-wide experiment), it is widely established in the state of Berlin, for example.

1.2.2 Principles of Tracking in *Hessen*

As illustrated above, traditional secondary schools and the two year comprehensive orientation stages co-exist in *Hessen*. As a further alternative, the institution of the *Gesamtschule* offers fully comprehensive education from grade 5-10. The exact wording of the school law regulation on tracking in *Hessen* is given in the Appendix to Chapter 1. In principle, after fourth grade, parents decide on the secondary school type of their children based on children's abilities and previous school performance. Parents may opt for the 'support stage' or a comprehensive school (*Gesamtschule*) in order to give their children more time to assess their abilities and interests. Especially, parents wishing that their children attend the higher secondary track (*Gymnasium*) but are not sure that they will be able to cope with the demands of this school type may make them join a 'support stage' or a comprehensive school. The distance between a pupils' place of residence and the location of the respective school is a further determinant that is known to drive the decision to attend a 'support stage' school *vs.* a tracked secondary school in fifth grade.[13] Some regions in *Hessen* do not offer 'support stages' so that children hardly have the choice to attend this school type.[14] However, the school law states that if the desired school type is not offered in a pupil's region of residence

[13] This is illustrated in Hessisches Kultusministerium (1995), p. 36.

[14] For example the city of Darmstadt offers no 'support stages' but those located at generally comprehensive schools.

the pupil has the right to attend this school type in another region (*cf. § 70*, school law of *Hessen*).

If the 'support stage' is chosen after fourth grade a decision on the final secondary track must be reached after sixth grade. Again, the parents have the primary authority to decide on the school type. However, if the desired track is the highest secondary school, selection to this school type depends on the 'support stage' teachers' approval.

1.2.3 Data Sets and Descriptive Analysis

This section presents some descriptive evidence indicating the quantitative dimension of the different tracking regimes and the streaming of pupils to the different secondary school types in *Hessen*. Further descriptive illustrations refer to the incidences of track modification and grade repetition[15] after pupils have been tracked by one or the other regime. Due to the preselection of different groups of pupils into the tracking regimes it is important to keep in mind that the presented stylized facts do not provide insights into the causal educational effects of one tracking regime compared to the other.

The following descriptive statistics are based on newly available individual level data provided by the local statistical office of the state of *Hessen*. The data set covers all pupils enrolled in general schools in *Hessen* in the school years 2002/2003 - 2005/2006. At time of writing this paper, besides the official statistical tables, there exist only two empirical studies drawing on this data base (Puhani and Weber, 2007a and Puhani and Weber, 2007b). One drawback of the data is that it does not provide a panel, *i.e.* pupils cannot be tracked using an individual identification number. Thus, even if several data waves exist, my analysis is based on a cross-section of observations. Little information is given on the prior development of the pupils (*i.e.* prior grade and school type) and this only refers to the previous year.

While the advantage of the data set is its large number of observations, a clear general disadvantage is the limited number of reported variables for each individual. Besides variables indicating region, school and class, individual information is given on gender, birth year and month, school entry year and month, and nationality. There are no outcome variables such as school marks or test scores. However, it is possible to identify the incidences of grade repetition and track modification (*i.e.* the correction of initial track choice) from one year to the following year.

The results based on the *Hessen* data are complemented by evidence based on the national PISA-E database covering about 2,300 ninth graders in the German state of *Hessen*.

[15] In Germany, low performing pupils have to repeat a grade if they are not able to attain certain marks.

The PISA-E data are a national extension of the international PISA 2000 data including supplementary questions from pupils and parents questionnaires as well as test results from the standardized math, reading and science tests. No information is available from school questionnaires which are included in the PISA study. The main reason why I use PISA-E instead of PISA is that information on 'support stage' attendance in fifth grade is only available in the extension study. Compared to the *Hessen* data, the advantage of PISA-E is that it allows to control for a variety of individual background characteristics. This is why the econometric part of this study (Section 1.3) focuses on the PISA-E data.

Table 1.2 and Table 1.3 detail the provision of different school types in *Hessen* based on the most recent wave of the administrative data-set: Generally, nearly 13 % of all the primary and secondary schools in *Hessen* offer 'support stages' (206 out of 1,642 schools as can be deduced from Table 1.2). Considering all schools offering secondary programmes, 15 % (87 out of 585) have fully comprehensive education programmes where pupils are not tracked into 'classical' secondary categories. Most of the 'support stages' are found at these fully comprehensive schools (45 % or 93 out of 206 schools). The remaining 'support stages' are located at elementary schools (22 %), schools hosting elementary schools as well as lower and intermediate secondary schools (17 %), schools offering the lower and intermediate secondary tracks (10 %) and schools offering elementary and lower secondary education (5 %). One further school offers elementary as well as intermediate education and hosts a 'support stage'. The corresponding numbers of pupils in each of these detailed primary and secondary types is provided in Table 1.3.

Table 1.4 considers the school track choice of pupils being streamed after fourth grade in 2003 and of those who opted for the 'support stage' in 2003 and are tracked after sixth grade (in 2005). The corresponding numbers are calculated using two different waves of the data so that both groups under examination consist of pupils from approximately the same cohorts. Results from Table 1.4 indicate that most of the fifth graders have already been tracked to the 'classical' secondary school levels: The majority of them attend the higher secondary track (38 %), while the intermediate and lower secondary levels are less popular (14 % and 5 % respectively). Furthermore, 15 % of all fifth graders attend fully comprehensive schools and 28 % opt for the 'support stages'. The latter group of pupils is mostly streamed to secondary levels after sixth grade (except of those 2 % who decide to attend fully comprehensive schools): Pupils tracked in seventh grade mostly enter the intermediate (46 %) or even the lower level (32 %) schools. There are no feasible gender differences when tracking to the secondary levels takes place after fourth grade. However, for the pupils tracked after the 'support stage', girls tend to choose higher educational tracks compared to their male classmates.

Additional evidence by nationality group is provided in Table 1.5. The two major subgroups under analysis are 'native' pupils (as defined by pupils holding nationalities of German-speaking countries) and pupils holding another nationality ('non-natives'). Furthermore, I look at the two most frequent immigrant groups, which refer to pupils holding Turkish (about 6 % of the considered fifth graders) or Italian and Greek nationalities (1.6 % of the sample).[16] I do not consider further nationality groups because of the smaller sample sizes of these groups.

While 'native' pupils are most often tracked to the highest secondary schools after fourth grade (41 %) a relatively small proportion of 'non-native' fifth graders attend these schools (19 % of all 'non-natives', only 13 % of pupils from Turkey and 18 % of pupils from Italy/Greece). Most pupils with an immigrant background opt for the 'support stages' (34 % of all 'non-natives', 38 % and 32 % for pupils from Turkey and Italy/Greece respectively). This is consistent with the idea that these schools give them more time to integrate and learn the German language before having to decide on their educational (and professional) future.

The educational decision after the 'support stages' differs between immigrants and natives as well: While the highest proportion of natives reaches the intermediate secondary track after attending the 'support stages' (48 %), immigrants are most often selected to the lowest secondary schools (49 % of all 'non-natives', even 53 % of pupils from Turkey and 54 % of pupils from Italy/Greece).

Table 1.5 already gives a hint that there is selection to the 'support stages' (according to immigrant background). Direct evidence on selection is provided in Table 1.6 and Table 1.7 based on the administrative data for *Hessen* and the PISA-E data respectively. The probit regression results suggest that pupils with a less advantaged socio-economic background are selected into the 'support stages'. Similar to the results presented in Table 1.5, Table 1.6 shows that immigrant children (and especially immigrants from Turkey and Italy and Greece) are more likely to opt for the 'support stages' than native pupils. However, the coefficient on general immigrant background turns insignificant if more control variables are added in Table 1.7. Specifically, parental background (which is probably correlated to immigrant background) seems to influence the decision: Children are especially less likely to attend the 'support stages' if their father holds a tertiary level degree but also if their father holds no vocational degree. Additionally, Table 1.6 shows that the county of residence significantly deter-

[16] The data at hand do not allow distinguishing between Greek and Italian nationals.

mines the probability to attend the 'support stage'. This is probably driven by the fact that the provision of 'support stages' varies between school districts. Based on the *Hessen* data, the results also suggest that male pupils and older pupils (*i.e.* those born after the end of June, which is the official cut-off date for school entry) are more likely to attend the 'support stages'. However, the corresponding coefficients turn insignificant in Table 1.7 (the birth-date effect is still 'marginally' significant at the 13%-level of significance).

Table 1.8 and Table 1.9 aim at answering the question whether modification of the initial track choice and grade repetitions are unusual if pupils are tracked after six instead of four years of comprehensive schooling. As described above, one rationale behind the 'support stages' is that children are given more time to develop their abilities and skills and to obtain more information on their educational performance before deciding on the secondary track. If it is true that tracking after sixth grade is based on more reliable information on the pupils' abilities, one would expect that *ex-post* modification of the initially chosen track and grade repetitions are not frequent under the later tracking regime.

Thus, Table 1.8 shows the proportions of pupils staying in the chosen track in fifth, sixth and seventh grade. As explained in Section 1.2, it is generally possible to modify the initially chosen track at any grade level, whilst track modification is somewhat complicated by different curricula at different school types. Note, that the data at hand are not available as a panel. Thus, it is principally not possible to observe individuals over time in order to determine whether the track modification behaviour of former 'support stage' pupils differs from other pupils. However, I use information on the *shares* of former 'support stage' pupils being in the respective school at a given grade level. Table 1.8 distinguishes between schools having no incoming pupils from 'support stages' in grade seven and those having high shares (80 % or more) of incoming 'support stage' pupils. Since the number of incoming 'support stage' pupils differs by school track, I additionally distinguish between school tracks.

Generally, for the schools not educating any former 'support stage' pupils, the proportion of pupils staying in the previously chosen school type when moving to the following grade after a given grade amounts to 98 % in grades five, six, and seven. The proportion of stayers is lower (96 %) in the seventh grade for schools primarily recruiting former 'support stage' pupils. The difference in the proportion of stayers between schools not educating any 'support stage' pupils and schools primarily educating 'support stage' pupils is especially high in the highest secondary school track: While 99 % of the seventh graders remain in the highest level school track in the schools without former 'support stage' pupils, only 94 % are stayers in the schools featuring a high proportion of former 'support stage' pupils. Even if one takes into account that the seventh graders in the first type of schools (no ‘support stage’ pupils) possibly already revised their initial track decision after grades five and six, the figure of six percent of track changers in the second type of schools (featuring a high share of ‘support stagers’) is comparably high.

All in all, a relatively high proportion of pupils in the higher secondary track decide to revise the track decision made after the ‘support stages’. While a primary objective of the ‘support stages’ is the optimisation of school track choice through a longer period of observation and support in the comprehensive system, the changer rates following the tracking grade suggest that the ‘support stage’ based decisions may not be as appropriate as expected. However, it must be noted again that this descriptive evidence does not provide causal effects of the tracking regime in the statistical sense but merely looks at the educational performance of self-selected groups of pupils who have chosen one or the other tracking regime.

Table 1.9 additionally presents proportions of grade retainees (*i.e.* pupils who have to repeat a grade due to poor performance) following the same strategy as Table 1.8 above. A casual examination of the first set of rows in Table 1.9 gives the impression that the proportion of pupils not succeeding in the given grade is especially high for schools with high shares of incoming ‘support stage’ pupils. However, if the proportion of retained pupils is calculated by school track type (see the next sets of rows in Table 1.9) it is shown that the high proportion of retainees in schools receiving high shares of former ‘support stage’ pupils is due to the fact that these schools are mainly at the lower or intermediate secondary level. There are no feasible differences in the proportions of retained pupils if the comparison relates to schools of the same track type.

1.3 Econometric Strategies and Regression Results

1.3.1 Identification Strategy and Specifications for the Econometric Analysis

If the tracking regime were randomly assigned, the causal effect of ‘support stage’ attendance on educational outcomes could be estimated using a simple OLS regression framework. The corresponding regression equation is given by:

$$Y_i^t = \beta X_i + \gamma S_i + \varepsilon_i \,, \qquad (1)$$

where Y_i^t is the educational outcome of individual *i* measured at time *t* (several years after the regime choice), X_i is a vector of explanatory variables, S_i refers to the tracking regime indicator, and ε_i is the error term. However, as stated above, the prior choice of the tracking regime is endogenous to educational outcomes. One may assume that pupils choosing to attend the 'support stages' differ from the average pupil in (unobserved) characteristics which are also related to the schooling outcome so that $corr(S_i, \varepsilon_i) \neq 0$. Thus, simply estimating the effect of 'support stage' attendance on later educational outcomes by OLS will yield biased results.

One standard solution to such an endogeneity problem is to apply an instrumental variable strategy. The crux is whether it is possible to find a valid instrument which explains 'support stage' attendance but is not correlated to unobservable characteristics driving the outcome variable. In my opinion, it is not possible to find a valid instrument.[17] Therefore, I opt for a different strategy to pin down the effect of 'support stage' attendance. Formally, I assume that the true model equation is:

$$Y_i^t = \beta X_i + \gamma S_i + \delta U_i, \qquad (2)$$

where U_i refers to a vector of non-controlled variables determining both the tracking regime choice after fourth grade and educational outcomes at a later point in time. The corresponding estimation equation is:

$$Y_i^t = \beta X_i + \gamma S_i + \delta U_i + u_i, \qquad (3)$$

where $corr(S_i, u_i) = 0$. Thus, the underlying problem is taken to be an omitted variable problem where the error term in equation 1 contains both the influences of the characteristics (δU_i) and the error term of equation 3 (u_i). The feasible solution to this problem is to control for as many of the variables (U_i) causing the bias as possible using a relatively rich data set (the PISA-E data) on the pupils' individual and family background.

Table 1.10 gives an overview of the different specifications used in the regression analysis. The variables covered by the different specifications are explained in more detail in Table 1.11. Specification 1 simply includes the dummy variable of interest (indicating whether the pupils attended the 'support stage' regime) and a control dummy variable for attending the fully comprehensive system. In other words: the regression results differentiate between effects of three options of tracking regimes (*i.e.* the earlier and the later tracking regime and the comprehensive system). Individual characteristics (gender, immigration background and a proxy for school entry age) are added in specification 2. Specification 3 addi-

[17] One potential instrument that springs to mind is the density of 'support stages' in a region: Using this instrument it is assumed that pupils are more likely to decide to opt for the 'support stage' regime if there are many 'support stage' schools in their county of residence. However, the provision of 'support stages' cannot be considered as exogenous to educational outcomes: The local 'support stage' density is potentially driven by the same or similar characteristics of a region's residents as the individual decision to attend the 'support stage'. Conducting regressions on the local provision of 'support stages' using county data shows that the local 'support stage' density is significantly determined by observable regional variables which are also thought to be important determinants of educational outcomes (for example income and wealth variables).

tionally includes family background variables (*i.e.* indicating the presence of parents at home, parental employment, education, and behavior and the presence of siblings). I assume that the endogeneity bias is reduced as one moves from specification 1 to specification 3. Especially, the variables added in specification 3 are mainly parental characteristics that influence the tracking regime choice as well as the children's educational outcomes. Ideally one would also directly control for initial ability of pupils, *i.e.* compare pupils who performed similarly before entering the different tracking systems. However, no appropriate performance measure is available in the data.[18]

A further issue is that in the PISA-E data there are missing observations for the variables of interest for some pupils. For each of the control variables up to five percent of the observations are missing. For parental education even 12 % (mother) and 16 % (father) of the observations are generally missing. Given that this might additionally bias the results, in the following regression analysis, I include dummy variables indicating missing observations.

In order to measure test results I use the averages of the plausible values of test scores which are given in PISA-E. For detailed information on the scaling of the PISA test results and test contents I refer the reader to the technical reports and documentaries (Adams and Wu, 2002 and especially the publication by Deutsches PISA Konsortium, 2003 for the German extension study). The plausible values correspond to the ones measured in the PISA-study but are standardized for each German state so that the mean score equals 100 and the standard deviation is 30 for each state. Thus, comparisons of test results across German states are not possible and analyses must be conducted at the single state's level.[19] For the sake of representativeness, all statistics are weighted using the sampling weights provided in the dataset.

The simple regressions only identify the impact of later tracking at the mean of the conditional performance distribution. As mentioned in the introduction, from a theoretical point of view, there are counteracting effects of later tracking: While later tracking may result in a more appropriate tracking decision because of improved information concerning the children's ability, more able pupils may actually benefit from early tracking *e.g.* through positive peer effects. Thus, it is interesting to examine whether the later tracking effect differs for pupils with a different background and of different ability. Therefore, the presentation of regression results is complemented by sub-group analyses focusing on pupils' family background.

[18] The only potential measure is the school level the pupil had been recommended to attend after fourth grade. For pupils attending the 'support stages' the indicated level might also be the one recommended after sixth grade and thus be an outcome of 'support stage' attendance. This is why I do not use this information.
[19] In the original PISA study scores are standardized to an international mean 500 and standard deviation 100 which allows international comparisons.

Additionally, quantile regressions are conducted in order to directly consider pupils at different positions of the conditional distributions of test scores.

1.3.2 Regression Results

Table 1.12 shows the results of OLS regressions of test performance on tracking regime dummies and different sets of explanatory variables (as explained in Table 1.10).[20] Generally, all the estimated effects are negative if they are significant. This might indicate that the attendance of a comprehensive class in fifth grade reduces school performance in ninth grade but the negative coefficients might also be the result of a negative selection of pupils into the comprehensive regimes after fourth grade. Including individual control variables in specification 2 hardly changes the estimated effects compared to specification 1. However, if parental background is considered in specification 3, the estimated coefficients decrease notably and become insignificant in most cases (except for the significance of the 'support stage' coefficient in the science regression and the coefficient on the comprehensive school indicator in the math regression).

The decrease in the absolute size of the negative coefficients as one moves from specification 2 to specification 3 reflects the 'negative selection' to the comprehensive school systems, *i.e.* pupils with a less favourable family background select to these systems (compare Section 1.2). This finding corresponds to a situation where low performers at elementary school who are recommended to the lower level schools opt for the comprehensive system in order to get a 'second chance' to find out whether they still have the ability to attend the high (or intermediate) level track.

Furthermore, the low and mostly insignificant effects for specification 3 indicate that the choice of the tracking system does not matter at least for the math and reading outcomes of ninth graders. Even if the identification strategy does not allow for the identification of the true causal effect of the tracking regime, because of the negative selection into the comprehensive systems (as indicated by the change in coefficients between specification 2 and 3) there is no reason to believe that the presented coefficients suffer from a downward bias. Thus, it is reasonable to conclude that there is no *negative* effect of 'support stage' (or comprehensive school) attendance on fifth graders math (or science) and reading performance.

---

[20] In addition to the presented regressions, I also conducted regressions where I allowed for a more flexible form by interacting the 'support stage' dummy and the explanatory variables. However, hardly any of the interaction coefficients proved to be significant in the full specification. Alternatively, I consider effects for some socio-economic sub-groups which will be discussed below.

Table 1.13 to Table 1.18 repeat the regressions for different sub-samples characterised by gender, immigrant background and parental characteristics. Generally, analysis by each gender yields similar findings as for the whole sample with the main conclusion that the 'support stage effect' drops down (mostly insignificant) if the full set of controls is included. However, there are two notable exceptions: For male pupils the negative reading score effect decreases but remains significant at the ten percent level and (more importantly) the negative science score effect does not decrease at all as more controls are included. Still, the methodological framework of this paper does not allow identifying whether the persistent negative effect concerning the science score is due to education in the 'support stage' or due to a persistent selection bias caused by unobserved characteristics.

Considering pupils with and without immigrant background, the following pattern emerges: For natives the 'support stage' effects decrease but remain significant (at least at the ten percent level) as the full set of controls is included. For immigrants the effect is insignificant or becomes insignificant if measured by the math and science score respectively. However, the immigrant pupils' reading score effect becomes significantly *positive* when using specification 3. If it is assumed that there is negative selection of pupils to the 'support stages' this finding suggests that there must be a positive regime effect related to the reading scores. Consequently, the results could be interpreted as demonstrating that immigrant pupils benefit (at least as far as their language skills are concerned) from being educated in the later tracking regime.

However, it might be argued that this conclusion only holds if there is in fact negative selection of immigrant pupils to the 'support stages'. This assumption would not be valid if immigrant pupils with initially higher language skills (*i.e.* pupils who have spent longer time in Germany and use the German language at home) self-selected to the 'support stages'. In order to take this objection into account, I estimate the 'support stage' effect separately for different groups of immigrants. The considered groups are: (1) pupils who were born abroad (*i.e.* mostly first generation immigrants), (2) pupils born in Germany whose parents were born abroad (second generation immigrants), (3) pupils who use a foreign language at home, (4) first generation immigrants who use a foreign language at home, and (5) second generation immigrants speaking a foreign-language at home. It is reasonable to assume that initial reading performance is better for second generation immigrants compared to first generation immigrants and especially compared to first generation immigrants speaking a foreign language at home.

The respective mathematics, reading and science score results by immigrant sub-group are presented in Table 1.15 - Table 1.17. Most of the findings considered are insignificant which might be due to limited sample sizes when considering sub-groups. However, looking at the point estimates, familiar patterns emerge for all sub-groups and subjects: If the ‘support stage’ effect is negative in the initial specification (without control variables) it decreases in absolute size or turns insignificant or positive in the full specification. For some sub-groups (second generation immigrants when considering mathematics; first generation immigrants and first generation immigrants using a foreign language at home for reading) the ‘support stage’ effect is positive even if no control variables are included. In these cases, the positive effect becomes more pronounced (and is significant for the reading score) if the full set of control variables is included. Interestingly, the positive ‘support stage’ effect in reading is especially high for first generation immigrants and first generation immigrants using a foreign language at home who might be considered to be a ‘negative selection’ (as concerns their initial reading skills) among the group of immigrant pupils. Since the positive effect becomes more pronounced as additional control variables are included, this is indicative of a negative selection bias being reduced. Summing up, I interpret these robust and consistent finding as supportive for the conclusion that ‘support stages’ are beneficial for the reading performance of immigrants.

Sub-group results by parental background are presented in Table 1.18. The considered groups are: (1) Children whose both parents are not employed, (2) children whose both parents do not hold a vocational degree, (3) children with a general ‘disadvantaged’ family background (*i.e.* children having either an immigrant background or having low educated or unemployed parents) and (4) children with an ‘advantaged’ family background (*i.e.* children having no immigrant background, no unemployed parent and no lowly educated parent). Since sample sizes drop to very small numbers for most of the sub-groups, I only present the results for the reading sample which is the largest sample. As a matter of fact, due to the limited sample size most of the sub-group results for the mathematics and science samples are insignificant (not shown here) but the general pattern emerging from these samples corresponds to the findings from the reading sample. The numbers of observations are already very limited for the reading regressions as can be deduced from Table 1.18. However, the results provide some interesting insights: First of all, and similar to Table 1.15 - Table 1.17 the ‘support stage’ effects are generally positive for the full specification when groups with a ‘disadvantaged’ family background are considered (in the first three columns of Table 1.18). These positive effects are significant or marginally significant (at the 10.5% level of significance in the third column). However, if children with a favourable family background are examined, the point estimate turns negative and is insignificant in the full specification. Thus, it seems that later tracking exerts different effects on different groups of children. If it is true that children with a less favourable family background benefit from the ‘support stages’ while this institution does not harm pupils with an advantaged family background, as it is suggested by these results, ‘support stages’ might reduce education inequality.

Distributional considerations are directly addressed using quantile regressions (Table 1.19). Figure 1.1 – Figure 1.3 show the estimated ‘support stage’ effects for different quantiles of the conditional test score distributions together with the mean regression results and its confidence bounds. An interesting pattern emerges for all test scores: While there are significant positive ‘support stage’ effects for the lower quantiles, the effect decreases nearly monotonically and turns to a significant positive effect for the upper quantiles. For the 10%-quantile for example the positive effect ranges between 5.35 scores for science and 6.65 for the reading score; this is equivalent to about one-fifth of the PISA-E standard deviation in the sample for *Hessen*. Looking at the 90%-quantile, the effect is also sizeable and ranges between -4.58 (science) and -4.14 (reading) which corresponds to about 15% of a standard deviation.

Thus, the quantile regression results suggest that ‘support stages’ work in favor of children with a disadvantaged education background whilst there are negative effects on pupils on top of the conditional performance distribution. Therefore, ‘support stages’ might reduce education inequality to the detriment of pupils on top of the (conditional) performance distribution. These findings are consistent with results from studies comparing tracking systems for different countries concluding that later tracking reduces education inequality (compare Section 1.1). Additionally, the theoretical literature on tracking provides explanations for the fact that tracking exerts differential impacts on pupils of different abilities: For example non-linear peer effects imply that high ability pupils specifically benefit from early segregation.

1.4 Conclusions of Chapter 1

The optimal tracking system is an issue of controversial discussion among educationalists and social scientists. This paper considered an alternative tracking regime which allows streaming pupils to secondary school types after six instead of four years in the German state of *Hessen*. It has been argued that pre-selection into the alternative tracking regime (*i.e.* the ‘support stages’) is not random. It seems that especially lower performers are selected to the later tracking regime. Thus, it is not surprising, that children attending the ‘support stages’ are more often tracked to the lower secondary school types later, as can be seen from the descriptive statistics.

In an attempt to reduce the endogeneity bias in estimating the regime choice effect, I controlled for a variety of individual and family characteristics such as parental education, employment and behavior. Overall, the estimated negative coefficients on the ‘support stage’ or comprehensive school indicators drop in absolute size as one controls for family background (and turn insignificant in most cases): I conclude that there seems to be no general negative effect of ‘support stage’ (or comprehensive school) attendance on educational outcomes of ninth graders when estimated at the mean. However, sub-group analyses reveal that later tracking exerts positive effects on pupils with a less favourable family background. The sub-group results are complemented by quantile regressions demonstrating that the ‘support stage’ effects decrease nearly monotonically over the conditional performance distributions. Thus, pupils at the lower quantiles benefit from later tracking in the sense that their PISA-E mathematics, reading and science score increase by one-fifth of a standard deviation. Unfortunately, the results also suggest that education inequality decreases to the detriment of ‘top performers’.

Recently, policy-makers in Germany discuss the modification of the tracking system. Whether another system is considered to be beneficial depends from the objectives behind such a reform. If the major objective is to improve the educational situation of ‘disadvantaged’ pupils and to reduce education inequality, evidence from this paper suggests that delaying the timing of tracking is favourable. However, one needs to bear in mind that such a reform might negatively impact the ‘top performers’.

For a more subtle analysis of causal effects of the later tracking regime more extensive individual level (panel) data sets on school attendance and performance are required. For future research, the recent changes in *Bremen* and *Niedersachsen* described in Section 1.2 may provide an interesting exogenous source of variation. The effects of these regime changes away from the later tracking system can probably be examined as soon as data on secondary educational performance of the cohorts affected by the regime change exists, given that it is made available to empirical researchers.

References for Chapter 1

Adams R. and M. Wu (2002): *PISA 2000. Technical Report*, OECD, Paris.

Ammermüller, A (2005): *Educational Opportunities and the Role of Institutions*, ZEW Discussion Paper 05-44.

Ariga, K., G. Brunello, R. Iwahashi and L. Rocco (2005): *Why Is the Timing of School Tracking So Heterogeneous?* IZA Discussion Paper 1854.

Bauer, P. and R.T. Riphahn (2006): Timing of School Tracking as Determinant of Intergenerational Transmission of Education, *Economic Letters* 91: 90-97.

Baumert, J., U. Trautwein and C. Artelt (2003): Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens, in Deutsches PISA-Konsortium (eds): *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Verlag Leske + Budrich, Opladen, 261-331.

Betts, J.R. and J.L. Shkolnik (2000): The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools, *Economics of Education Review* 19: 1-15.

Brunello, G., M. Giannini and K. Ariga (2007): The Optimal Timing of School Tracking, in P. Peterson and L. Wößmann (eds.): *Schools and the Equal Opportunity Problem,* MIT Press, Cambridge USA (forthcoming).

Deutsches PISA Konsortium (2003): *Ergänzungen zum Skalenhandbuch*, Max-Planck-Institut für Bildungsforschung, Berlin.

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

Eiko, J. (1991): *20 Jahre Orientierungsstufe: Beiträge zu einer umstrittenen Schulreform*, Academia Verlag, Sankt Augustin.

Eiko, J. (1989): *Die Orientierungsstufe im Urteil von Eltern und Lehrern*, Haag und Herchen, Frankfurt am Main.

Entorf, H. and M. Lauk (2006): *Peer Effects, Social Multipliers and Migrants at School: An International Comparison*, Darmstadt Discussion Papers in Economics 164.

Epple, D., E. Newton and R. Romano (2002): Ability Tracking, School Competition, and the Distribution of Educational Benefits, *Journal of Public Economics* 83: 1-48.

Fernandez, R. (1998): *Education and Borrowing Constraints: Tests vs. Prices*, Starr Center for Applied Economics Economic Research Reports 98-17.

Figlio, D.N. and M.E. Page (2002): School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?, *Journal of Urban Economics* 51: 497-514.

Hanushek, E.A. and L. Wößmann (2006): Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries, *The Economic Journal* 116: 63-76.

Henze G., U. Sandfuchs and C. Zumhasch (1996): *Fördern in der Orientierungsstufe: Eine empirische Untersuchung zur Förderung lernschwacher Schüler in Kleingruppen*, Klinkhardt Verlag, Bad Heilbrunn.

Hessisches Kultusministerium (1995): *Empfehlungen für die Arbeit in der Förderstufe*, Hessisches Institut für Bildungsplanung und Schulentwicklung, Wiesbaden.

Hopf, V. (1979): *Orientierungsstufe als Förderstufe*, Diesterweg, Frankfurt am Main.

Jürgens, E. (1991): *20 Jahre Orientierungsstufe. Beiträge zu einer umstrittenen Schulform*, Academia Verlag, Sankt Augustin.

OECD (2004): *Learning for Tomorrow's World – First Results from PISA 2003*, OECD, Paris.

Pekkarinen, T. (2005): *Gender Differences in Educational Attainment: Evidence on the Role of the Tracking Age from a Finnish Quasi-Experiment*, IZA Discussion Paper 1897.

Pischke, J.S. (2003): *The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years*, NBER Working Paper 9964.

Puhani, P.A. and A.M. Weber (2007a): Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany, *Empirical Economics* 32:359-386.

Puhani, P.A. and A.M. Weber (2007b): *How Persistent Is the Age of School Entry Effect in a System of Flexible Tracking? Regression Discontinuity Estimates Using Administrative Data on All Students in the German State of Hessen*, working paper, mimeo.

Schütz, G., H.W. Ursprung and L. Wößmann (2005): *Education Policy and Equality of Opportunity*, IZA Discussion Paper 1906.

Zimmer, R. (2003): A New Twist in the Educational Tracking Debate, *Economics of Education Review* 22: 307-315.

Tables and Figures for Chapter 1

Table 1.1: First age of selection in the education system

10	11	12	13	14	15	16
Austria	Czech Republic	Belgium	Canada	Italy	France	Australia
Germany	Hungary	Mexico	Luxembourg	Korea	Greece	Denmark
	Slovak Republic	Netherlands			Ireland	Finland
	Turkey				Japan	Iceland
					Poland	New Zealand
					Portugal	Norway
					Switzerland	Spain
						Sweden
						U.K.
						U.S.

Source: OECD (2004), page 262.

Table 1.2: Frequencies of primary and secondary school types in *Hessen*:

Detailed type of school (offered programmes)	absolute frequency	(%)	support stages	integrated comprehensive
elementary school	1,057	(64.37)	45	0
elementary + lower secondary	47	(2.86)	10	0
elementary + intermediate secondary	2	(0.12)	1	0
elementary + lower/intermediate secondary	65	(3.96)	36	0
lower secondary	6	(0.37)	0	0
intermediate secondary	25	(1.52)	0	0
lower + intermediate secondary	62	(3.78)	21	0
higher secondary	143	(8.71)	0	0
further combined elementary + secondary	39	(2.38)	10	11
further combined secondary	196	(11.94)	83	76

Note: + indicates that several school types are located in the same school building or area. This does not necessarily mean that school types offer ‘integrated’ (comprehensive) education. The numbers are calculated using the school-ID numbers in the data-set and considering for each school (as identified by its ID-number) the school types reported for the pupils of this school.
Source: Administrative data for *Hessen*, wave 2005/2006.

Table 1.3: Numbers of pupils in different primary and secondary school types in *Hessen*:

Detailed type of school (offered programmes)	absolute frequency	(%)	in support stages	in integrated comprehensive
elementary school	221,303	(32.86)	3,153	0
elementary + lower secondary	15,850	(2.35)	607	0
elementary + intermediate secondary	245	(0.04)	22	0
elementary + lower/intermediate secondary	38,753	(5.75)	4,169	0
lower secondary	1,493	(0.22)	0	0
intermediate secondary	12,917	(1.92)	0	0
lower + intermediate secondary	37,992	(5.64)	3,585	0
higher secondary	142,196	(21.12)	0	0
further combined elementary + secondary	27,245	(4.05)	1,464	4,755
further combined secondary	175,392	(26.05)	16,149	53,284

Note: + indicates that several school types are located in the same school building or area. This does not necessarily mean that school types offer 'integrated' (comprehensive) education. The numbers are calculated using the school-ID numbers in the data-set and considering for each school (as identified by its ID-number) the school types reported for the pupils of this school.
Source: Administrative data for *Hessen*, wave 2005/2006.

Table 1.4: Track choice in the earlier and in the later tracking regime

Selection after / into	4th grade (tracking of *all* pupils)			6th grade (tracking of *support stage* pupils)		
	all (%)	male (%)	female (%)	all (%)	male (%)	female (%)
lower secondary	4.64	5.13	4.14	32.09	35.49	28.42
intermediate secondary	14.40	14.39	14.42	46.37	44.66	48.22
higher secondary	37.74	36.16	39.37	19.15	17.24	21.21
fully comprehensive	15.27	15.59	14.95	2.38	2.61	2.14
support stage	27.94	28.73	27.13	---	---	---

Note: Sample of all pupils tracked after fourth grade of elementary school in 2003/2004 and after sixth grade of the 'support stage' in 2005/2006 respectively.
Source: Administrative data for *Hessen*, wave 2003/2004 and 2005/2006, own calculations.

Table 1.5: Track choice by nationality

Selection after / into	4th grade (tracking of *all* pupils)				6th grade (tracking of *support stage* pupils)			
	native	non-native	Turkey	Italy/ Greece	native	non-native	Turkey	Italy/ Greece
lower secondary	3.66	10.53	10.78	11.25	28.65	49.26	52.64	53.57
intermediate sec.	13.74	18.38	20.05	17.19	47.99	38.31	37.12	37.14
higher secondary	40.96	18.56	13.00	18.02	20.99	9.97	7.54	6.79
comprehensive	14.69	18.72	18.59	21.67	2.37	2.45	2.71	2.50
support stage	29.96	33.81	37.58	31.87	---	---	---	---

Note: Sample of all pupils tracked after fourth grade of elementary school in 2003/2004 and after sixth grade of the 'support stage' in 2005/2006 respectively.
Source: Administrative data for *Hessen*, wave 2003/2004 and 2005/2006, own calculations.

Table 1.6: Probit regressions on the selection to 'support stages' (*Hessen* data)

Variables	marginal effects	(s.e.)
gender (male)	0.02**	(0.00)
born in July-December (proxy for late school entry)	0.02**	(0.00)
Nationality indicators (Reference: German speaking countries):		
Turkey	0.14**	(0.01)
Italy / Greece	0.11**	(0.02)
former Yugoslavia	0.06**	(0.02)
further Western countries	0.09**	(0.02)
further Eastern countries	0.06**	(0.02)
Muslim countries (without Turkey)	0.09**	(0.02)
remaining Asian countries	0.02	(0.02)
remaining countries	0.16**	(0.03)
Regional variables (Reference for indicator variables: Frankfurt):		
region 1: Darmstadt/Dieburg	-0.09**	(0.01)
region 2: Offenbach	0.34**	(0.02)
region 3: Wiesbaden	-0.12**	(0.01)
region 4: Bergstraße, Odenwald	-0.11**	(0.01)
region 5: Groß-Gerau	0.41**	(0.01)
region 6: Hochtaunus	0.00	(0.01)
region 7: Main-Kinzig	0.06**	(0.01)
region 8: Main-Taunus	-0.06**	(0.01)
region 9: Offenbach-Land	-0.01	(0.01)
region 10: Rheingau-Taunus	0.16**	(0.01)
region 11: Wetterau	-0.02	(0.01)
region 12: Gießen	0.17**	(0.01)
region 13: Lahn-Dill	0.28**	(0.01)
region 14: Limburg-Weilburg	0.13**	(0.01)
region 15: Marburg-Biedenkopf, Vogelsberg	0.06**	(0.01)
region 16: Kassel	0.39**	(0.01)
region 17: Fulda, Hersfeld-Rotenburg	0.19**	(0.01)
region 18: Kassel-Land, Werra-Meißner	0.19**	(0.01)
region 19: Schwalm-Eder, Waldeck-Frankenberg	0.40**	(0.01)
region 20: Frankfurt	0.26**	(0.01)
observations	63,888	
# support stage	18,342	

Note: The regression is based on fifth graders in 2003/2004. Marginal effects are reported. For dummy variables, the marginal effects correspond to a change from 0 to 1. All control variables available in the data set are included. * Significant at the ten percent level. ** Significant at the five percent level.

Source: Administrative data for *Hessen*, wave 2003/2004, own estimations.

Table 1.7: Probit regressions on the selection to 'support stages' (PISA-E data)

Variables	marginal effects	(s.e.)
gender (male)	-0.03	(0.02)
immigrant background	-0.03	(0.02)
born in July-December (proxy for late school entry)	0.03	(0.02)
father lives at home	-0.03	(0.06)
mother lives at home	-0.03	(0.03)
mother employed	-0.04	(0.02)
father employed	-0.03	(0.04)
mother: no vocational eduaction	0.04	(0.04)
mother: tertiary education	0.06	(0.04)
father: no vocational eduaction	-0.09**	(0.04)
father: tertiary education	-0.13**	(0.02)
parental reading encouragement	-0.05**	(0.03)
number of siblings	-0.01	(0.02)
number of siblings squared	0.00	(0.00)
observations	2,334	
# support stage	468	

Note: The regression is based on ninth graders in the PISA-E 2000 reading sample indicating their 'support stage' attendance in fifth grade. Marginal effects are reported. For dummy variables, the marginal effects correspond to a change from 0 to 1. The control variables correspond to the variables used in Section 1.3 of this paper (see Table 1.11 for a detailed explanation of the variables). * Significant at the ten percent level. ** Significant at the five percent level.

Source: PISA-E 2000, own estimations.

Table 1.8: Proportions of stayers in school tracks by previous ‘support stage’ attendance

Stayers after …	No incoming support stage pupils (0%)			High share of incoming support stage pupils (>80%)		
	All Track Types					
	ratio	(s.d.)	observ.	ratio	(s.d.)	observ.
… 5th grade (2003/04)	0.98	(0.14)	15,938	---	---	---
… 6th grade (2004/05)	0.98	(0.13)	16,053	---	---	---
… 7th grade (2005/06)	0.98	(0.14)	15,937	0.96	(0.18)	13,877
	Lower Secondary					
… 5th grade (2003/04)	0.97	(0.17)	1,640	---	---	---
… 6th grade (2004/05)	0.96	(0.19)	1,859	---	---	---
… 7th grade (2005/06)	0.98	(0.13)	1,975	0.99	(0.11)	4,561
	Intermediate Secondary					
… 5th grade (2003/04)	0.95	(0.23)	3,539	---	---	---
… 6th grade (2004/05)	0.96	(0.21)	3,579	---	---	---
… 7th grade (2005/06)	0.95	(0.21)	3,620	0.96	(0.19)	6,455
	Higher Secondary					
… 5th grade (2003/04)	0.99	(0.09)	10,759	---	---	---
… 6th grade (2004/05)	0.99	(0.08)	10,615	---	---	---
… 7th grade (2005/06)	0.99	(0.10)	10,342	0.94	(0.24)	2,861

Note: The ‘proportions of stayers’ indicate the number of pupils in the given school type divided by the number of pupils in the given school type who have already been in this school the year before. Only pupils in tracked school types moving from one grade to the following grade (*e.g.* from grade 5 to grade 6 in 2003/2004) are considered. The total number of pupils in a given grade is not equal to the total number of pupils in the previous grade times the proportion of stayers since grade retainees additionally lower the number of remaining pupils. Pupils dropping out of the school system or moving to another German state are not observed, grade retainees are not considered. Proportions are separately calculated for schools with no incoming ‘support stage’ pupils and schools with high shares of incoming ‘support stage’ pupils. The share of incoming pupils from the ‘support stages’ is calculated by the proportion of seventh graders in the respective school in 2004/2005 having attended ‘support stages’ in sixth grade. The proportions are very similar (and thus robust) if grade retainees are kept in the sample.
Source: Administrative data for *Hessen*, waves 2003/2004 to 2005/2006, own calculations.

Table 1.9: Proportions of retained pupils by share of incoming 'support stage' pupils

Retainees in …	No incoming support stage pupils (0%)			High share of incoming support stage pupils (>80%)		
	All Track Types					
	retained	(s.d.)	observ.	retained	(s.d.)	observ.
… 5th grade (2003/04)	0.03	(0.17)	16,417	---	---	---
… 6th grade (2004/05)	0.03	(0.16)	16,480	---	---	---
… 7th grade (2005/06)	0.04	(0.20)	16,550	0.07	(0.25)	14,789
	Lower Secondary					
… 5th grade (2003/04)	0.07	(0.26)	1,765	---	---	---
… 6th grade (2004/05)	0.06	(0.23)	1,973	---	---	---
… 7th grade (2005/06)	0.08	(0.27)	2,261	0.08	(0.28)	5,070
	Intermediate Secondary					
… 5th grade (2003/04)	0.04	(0.20)	3,693	---	---	---
… 6th grade (2004/05)	0.04	(0.20)	3,736	---	---	---
… 7th grade (2005/06)	0.06	(0.24)	3,806	0.07	(0.26)	6,970
	Higher Secondary					
… 5th grade (2003/04)	0.02	(0.13)	10,959	---	---	---
… 6th grade (2004/05)	0.01	(0.12)	10,771	---	---	---
… 7th grade (2005/06)	0.03	(0.17)	10,483	0.03	(0.16)	2,749

Note: The 'proportions of retained pupils' indicate the number of pupils attending the same grade as in the previous year divided by the number of pupils at the given grade level. Only pupils in tracked school types are considered. Pupils dropping out of the school system or moving to another German state are not observed. Retainees include pupils changing to another track if they are repeating the grade in this track. Proportions are separately calculated for schools with no incoming 'support stage' pupils and schools with high shares of incoming 'support stage' pupils. The share of incoming pupils from the 'support stages' is calculated by the proportion of seventh graders in the respective school in 2004/2005 having attended 'support stages' in sixth grade.
Source: Administrative data for *Hessen*, waves 2003/2004 to 2005/2006, own calculations.

Table 1.10: Specifications for the econometric analysis

Specification	Included Variables
specification 1	tracking regime indicators
specification 2	specification 1 + individual characteristics (gender, immigration background indicator, proxy for school entry age)
specification 3	specification 2 + family background (presence of parents at home, employment of parents, education of parents, parental reading encouragement, siblings)

Note: The variables used in the different specifications are explained in Table 1.11.

Table 1.11: Variables used in the different specifications

Variable	**Explanation**
Tracking Regime Indicators (Reference = Tracking after fourth grade):	
support stage	dummy variable for 'support stage' attendance in fifth grade
comprehensive school	dummy for comprehensive school attendance in fifth grade
Variables Added in Specification 2 (Individual Characteristics):	
gender	dummy for male gender
immigration	dummy indicating whether pupil or parents were born abroad
proxy for school entry age	dummy indicating whether pupil is born before the official school entry cut-off date of June (= theoretically entered school relatively young according to the official school entry rule)[A]
Variables Added in Specification 3 (Family Characteristics):	
father	dummy indicating whether only a male guardian (mostly the father) lives with the child
mother	dummy indicating whether only a female guardian (mostly the mother) lives with the child
employment of mother	dummy indicating whether the mother is employed
employment of father	dummy indicating whether the father is employed
mother: no vocational education [B]	dummy indicating whether mother does not hold a vocational degree
mother: tertiary education [B]	dummy indicating whether mother holds a tertiary educational degree
father: no vocational education [C]	dummy indicating whether mother does not hold a vocational degree
father: tertiary education [C]	dummy indicating whether mother holds a tertiary educational degree
parental reading encouragement	parents often read to child before child learned to read
siblings	dummy indicating whether there are siblings of the child

Note: [A] See the paper by Puhani and Weber (2007a) or Chapter 2 for the motivation of this variable. [B] The reference category are mothers holding a vocational (upper secondary) degree. [C] The reference category are fathers holding a vocational (upper secondary) degree. In addition to these variables dummy variables for missing information are included.

Table 1.12: Results of OLS regressions of PISA-E scores on 'support stage' attendance

Test		Maths	Reading	Science
specifi-cation	regime	coefficients (s.e.)	coefficients (s.e.)	coefficients (s.e.)
1	support stage	-5.90** (2.39)	-4.12** (1.63)	-7.17** (2.54)
	comprehensive school	*-6.65** (2.37)*	*-2.67 (1.67)*	*0.71 (2.20)*
2	support stage	-5.38** (2.47)	-4.39** (1.63)	-8.48** (2.47)
	comprehensive school	*-7.24** (2.29)*	*-3.28** (1.61)*	*0.59 (1.10)*
3	support stage	-1.94 (2.14)	-1.08 (1.47)	-5.25** (2.29)
	comprehensive school	*-4.68** (2.10)*	*-0.96 (1.48)*	*2.45 (2.07)*
	observations	1,222	2,306	1,262
	# support stage in 5th grade	245	464	261
	# compr. school in 5th grade	208	386	196

Note: The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.13: Regression results by gender

Test		Maths		Reading		Science	
specifi-cation	regime	coefficients (s.e.)		coefficients (s.e.)		coefficients (s.e.)	
		female	male	female	male	female	male
SPECIFICATION 1	support stage	-6.04	-5.28*	-3.23	-5.63**	-6.11*	-7.95**
		(4.02)	(2.79)	(2.39)	(2.18)	(3.62)	(3.55)
	comprehensive	*-10.44***	*-2.91*	*-4.67***	*-1.27*	*0.35*	*1.14*
		(3.35)	*(3.15)*	*(2.25)*	*(2.45)*	*(2.75)*	*(3.33)*
2	support stage	-5.44	-5.95**	-3.37	-5.74**	-7.52**	-9.54**
		(4.24)	(2.66)	(2.46)	(2.16)	(3.59)	(3.41)
	comprehensive	*-10.89***	*-4.51*	*-4.61*	*-2.26*	*0.34*	*0.03*
		(3.50)	*(2.97)*	*(2.23)*	*(2.33)*	*(2.80)*	*(3.01)*
3	support stage	-0.32	-3.35	0.98	-3.43*	-2.48	-7.98**
		(3.13)	(2.60)	(2.08)	(2.06)	(3.18)	(3.02)
	comprehensive	*-7.06***	*-2.88*	*-1.49*	*-0.91*	*2.25*	*0.91*
		(3.02)	*(2.83)*	*(1.96)*	*(2.17)*	*(2.65)*	*(2.96)*
observations		548	674	1,074	1,232	577	685
# support stage		114	131	224	240	117	144
# comprehensive school		96	112	190	196	90	106

Note: The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.14: Regression results by immigration background[A]

Specification	Regime	Maths native	Maths immigrant	Reading native	Reading immigrant	Science native	Science immigrant
		coefficients (s.e.)		coefficients (s.e.)		coefficients (s.e.)	
1	support stage	-9.29**	-0.56	-6.70**	0.45	-7.63**	-11.61**
		(2.35)	(5.15)	(1.90)	(3.16)	(3.07)	(4.41)
	comprehensive	*-9.12***	*-4.96*	*-3.75***	*-2.19*	*0.14*	*1.53*
		(2.94)	*(3.77)*	*(1.91)*	*(2.99)*	*(2.65)*	*(3.43)*
2	support stage	-9.28**	0.20	-6.89**	0.58	-7.52**	-10.52**
		(2.30)	(4.97)	(1.86)	(3.04)	(2.97)	(4.38)
	comprehensive	*-8.92***	*-4.35*	*-3.74**	*-2.54*	*-0.18*	*1.46*
		(2.78)	*(3.95)*	*(1.92)*	*(2.95)*	*(2.57)*	*(3.39)*
3	support stage	-4.87**	2.59	-3.11*	4.19*	-4.89*	-5.12
		(2.32)	(3.84)	(1.77)	(2.47)	(2.74)	(3.59)
	comprehensive	*-6.96***	*0.42*	*-2.27*	*2.66*	*0.47*	*6.67**
		(2.39)	*(3.80)*	*(1.74)*	*(2.80)*	*(2.49)*	*(3.62)*
observations		802	420	1,562	744	866	396
# support stage		169	76	329	135	192	69
# comprehensive school		148	60	274	112	136	60

Note: 'Immigrant' refers to pupils who were born abroad or whose parents were born abroad (compare Table 1.11). The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.15: Mathematics regression results for different groups of immigrants

Specification	Regime	First generation immigrants	Second generation immigrants	Foreign language spoken at home	First generation immigrants + foreign language spoken at home	Second generation immigrants + foreign language spoken at home
1	support stage	-3.56	3.50	-5.10	-2.72	-12.69
		(4.68)	(9.66)	(4.16)	(4.76)	(7.83)
	comprehensive	*-1.86*	*-8.64**	*-2.27*	*-2.06*	*-9.42**
		(5.36)	*(5.19)*	*(4.21)*	*(5.56)*	*(5.12)*
2	support stage	-2.25	2.45	-4.92	-1.49	-15.79**
		(4.27)	(9.67)	(3.81)	(4.24)	(6.76)
	comprehensive	*-1.53*	*-8.03*	*-2.33*	*-1.26*	*-8.54*
		(5.66)	*(5.38)*	*(4.35)*	*(5.86)*	*(5.42)*
3	support stage	-1.38	5.65	-1.90	1.22	-11.75**
		(4.01)	(5.75)	(3.48)	(4.16)	(5.94)
	comprehensive	*0.97*	*2.86*	*1.35*	*1.45*	*0.82*
		(5.25)	*(6.55)*	*(3.98)*	*(5.46)*	*(7.37)*
observations		227	193	300	200	86
# support stage		48	28	50	37	12
# comprehens. school		31	29	51	29	16

Note: The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.16: Reading regression results for different groups of immigrants

Specification	Regime	First generation immigrants	Second generation immigrants	Foreign language spoken at home	First generation immigrants + foreign language spoken at home	Second generation immigrants + foreign language spoken at home
1	support stage	6.52*	-5.98	-1.89	4.29	-15.43**
		(3.45)	(6.08)	(3.05)	(3.51)	(6.02)
	comprehensive	*2.94*	*-7.50**	*-0.42*	*3.22*	*-5.96*
		(3.88)	*(4.53)*	*(3.27)*	*(4.08)*	*(6.07)*
2	support stage	6.62**	-6.18	-1.66	4.35	-15.67**
		(3.37)	(5.96)	(3.03)	(3.47)	(6.33)
	comprehensive	*2.45*	*-8.17**	*-0.49*	*2.79*	*-6.62*
		(3.89)	*(4.45)*	*(3.23)*	*(4.08)*	*(5.84)*
3	support stage	10.22**	-2.71	1.77	7.10**	-7.96
		(3.11)	(3.89)	(2.96)	(3.32)	(5.46)
	comprehensive	*5.45*	*0.76*	*2.61*	*4.70*	*0.63*
		(3.75)	*(4.07)*	*(3.18)*	*(3.86)*	*(5.16)*
observations		386	358	539	334	176
# support stage		82	53	101	68	28
# comprehens. school		53	59	88	48	32

Note: The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.17: Science regression results for different groups of immigrants

Specification	Regime	First generation immigrants	Second generation immigrants	Foreign language spoken at home	First generation immigrants + foreign language spoken at home	Second generation immigrants + foreign language spoken at home
1	support stage	-8.87	-13.83*	-10.89**	-9.29	-15.34
		(5.63)	(7.09)	(5.45)	(6.77)	(10.31)
	comprehensive	*3.76*	*-0.67*	*1.27*	*4.88*	*-3.86*
		(5.25)	*(4.44)*	*(4.13)*	*(5.79)*	*(5.76)*
2	support stage	-7.61	-13.23*	-9.94*	-8.06	-15.52
		(5.64)	(7.17)	(5.39)	(6.65)	(10.53)
	comprehensive	*3.01*	*-1.35*	*0.39*	*4.00*	*-7.01*
		(5.34)	*(4.54)*	*(4.08)*	*(5.80)*	*(5.57)*
3	support stage	1.52	-8.88	-5.12	0.08	-13.26
		(4.74)	(5.74)	(4.57)	(5.89)	(8.80)
	comprehensive	*9.64*	*2.77*	*5.57*	*9.07*	*-2.06*
		(6.09)	*(4.76)*	*(4.56)*	*(6.66)*	*(5.99)*
observations		203	193	286	174	98
# support stage		38	31	52	30	19
# comprehens. school		30	30	46	25	17

Note: The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.18: Reading regression results according to family background

	Regime		Both parents not working	Both parents low educated	Less favourable family background	Favourable family background
SPECIFICATION	1	support stage	6.20	13.95**	0.57	-5.25**
			(7.04)	(5.29)	(2.91)	(2.01)
		comprehensive	*3.16*	*11.30**	*-1.56*	*-2.88*
			(7.28)	*(6.66)*	*(2.72)*	*(2.09)*
	2	support stage	9.58	10.15*	-0.28	-5.25**
			(7.62)	(5.51)	(2.63)	(1.94)
		comprehensive	*1.26*	*5.39*	*-2.28*	*-2.71*
			(7.19)	*(6.96)*	*(2.72)*	*(2.12)*
	3	support stage	18.10**	10.85*	3.60	-2.89
			(7.36)	(6.15)	(2.23)	(1.83)
		comprehensive	*5.14*	*3.95*	*1.50*	*-1.88*
			(8.55)	*(7.32)*	*(2.62)*	*(2.03)*
observations			123	104	846	1,229
# support stage			26	78	714	962
# comprehens. school			19	26	132	267

Note: Results are only presented for the reading sample, because sample sizes are even smaller for the science and mathematics test. The reported coefficients refer to the 'support stage' dummy and the dummy variable for attendance of a general comprehensive school in fifth grade. The different specifications are explained in Table 1.10. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Table 1.19: Quantile regression results

Quantiles	Maths	Reading	Science
0.10	6.23**	6.65**	5.35**
	(2.11)	(2.24)	(2.36)
0.20	3.94*	2.03	2.92
	(2.21)	(2.32)	(2.11)
0.30	3.06	2.68	2.32
	(2.14)	(2.09)	(1.64)
0.40	2.67	2.48	1.74
	(1.78)	(1.92)	(2.03)
0.50	1.32	1.48	1.97
	(1.63)	(1.78)	(1.78)
0.60	-0.94	-0.71	0.35
	(1.65)	(1.79)	(1.65)
0.70	-2.57	-1.58	-1.39
	(1.86)	(1.83)	(1.94)
0.80	-3.73**	-3.68	-3.72*
	(1.77)	(2.05)	(1.94)
0.90	-4.25**	-4.14**	-4.58**
	(2.13)	(1.96)	(2.00)
0.99	-8.45**	-9.79**	-9.98**
	(2.17)	(2.36)	(2.32)

Note: The reported coefficients refer to the 'support stage' effect in the regressions using all control variables. Numbers in parentheses are the bootstrapped standard errors. The effects are also illustrated in Figure 1.1 – Figure 1.3. * Significant at the ten percent level. ** Significant at the five percent level.
Source: PISA-E 2000, own estimations.

Figure 1.1: 'Support stage' effects on PISA-E maths scores by quantiles

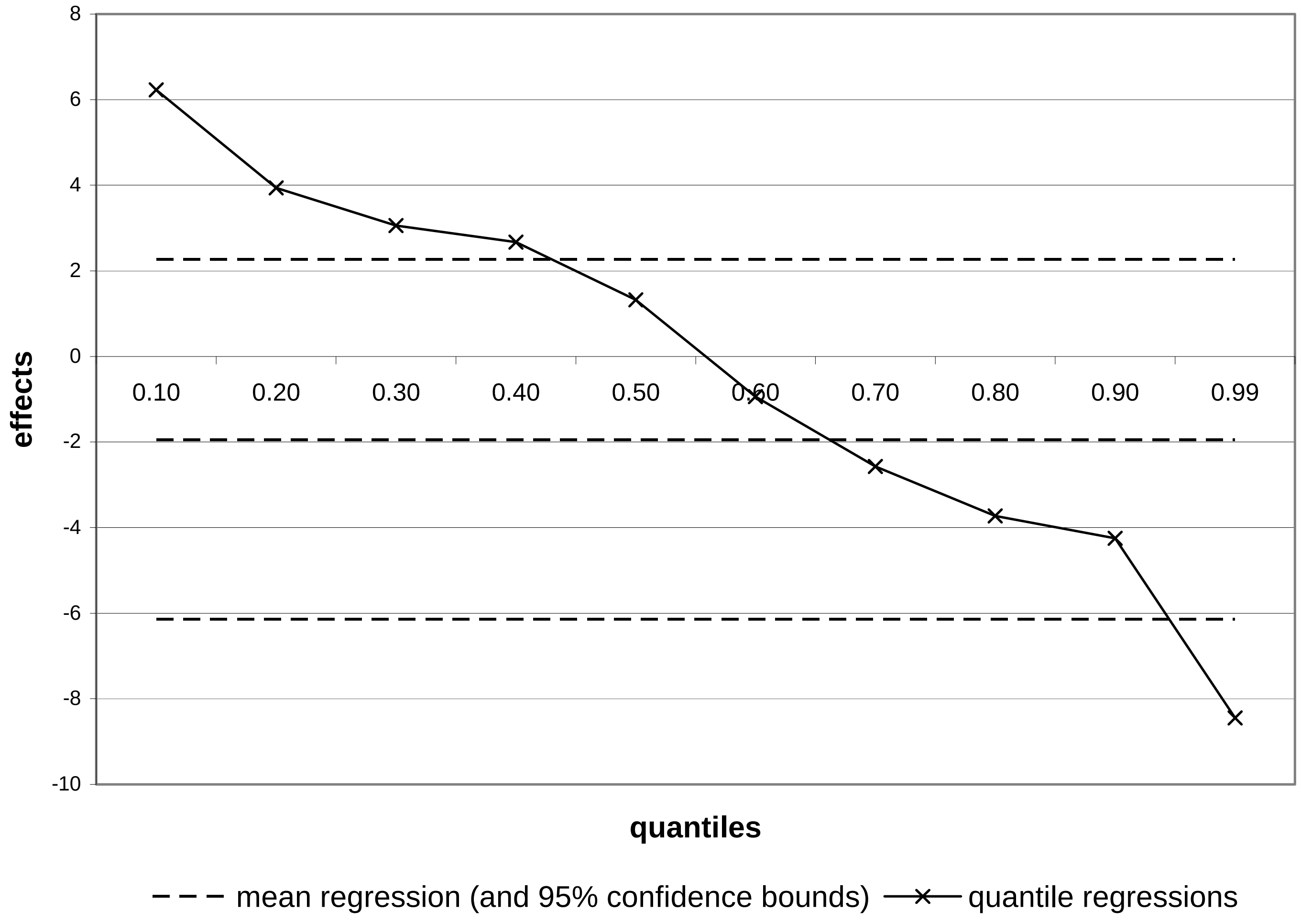


Figure 1.2: 'Support stage' effects on PISA-E reading scores by quantiles

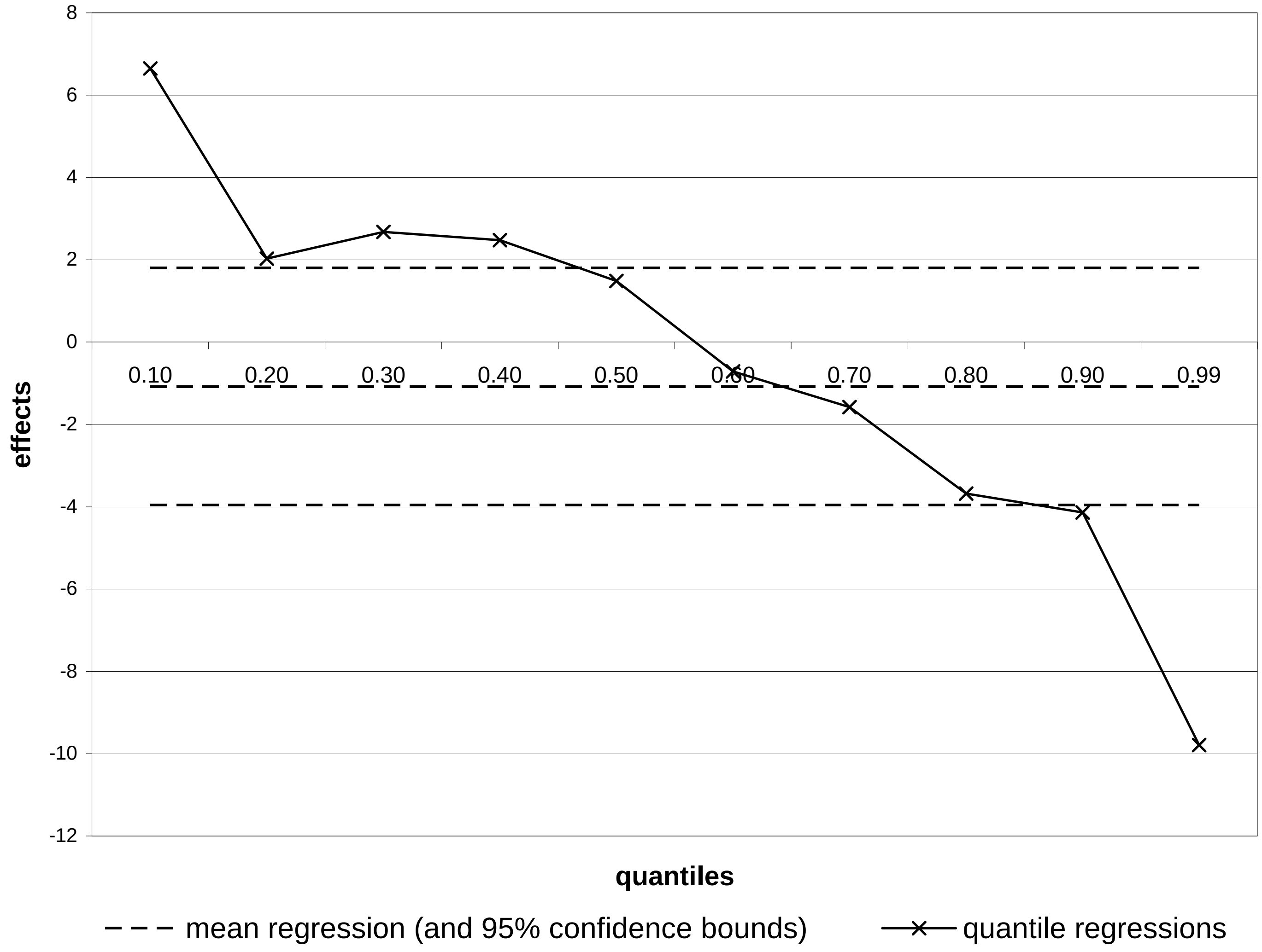
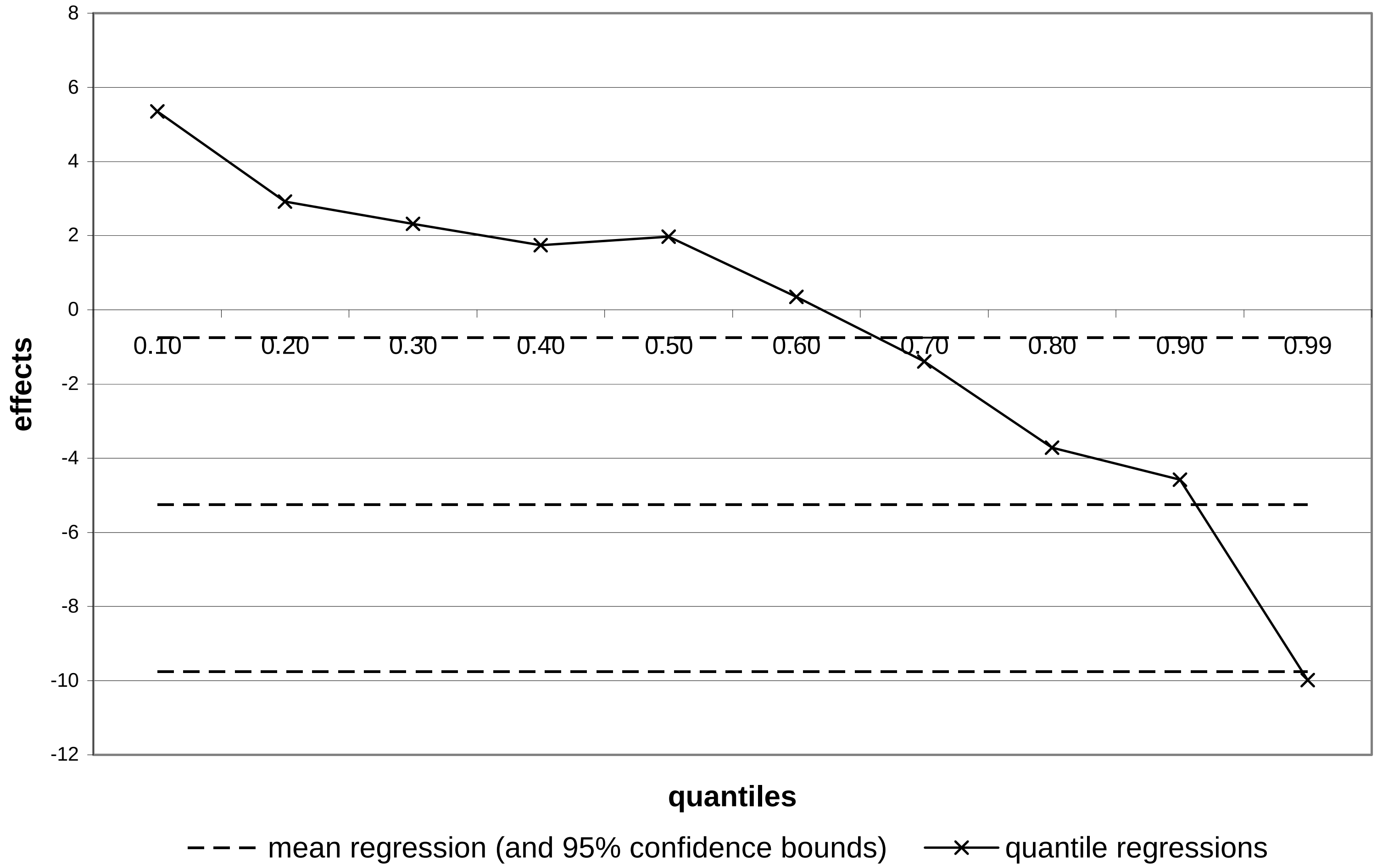


Figure 1.3: 'Support stage' effect on PISA-E science scores by quantiles



Appendix to Chapter 1 (Section 1.2.2)

Exact wording and English translation of main regulations on tracking in the school law of *Hessen* (*Hessisches Schulgesetz*), § 77:

(1) Die Wahl des Bildungsganges nach dem Besuch der Grundschule ist Sache der Eltern. Wird der Bildungsgang sowohl schulformbezogen als auch integriert angeboten, können die Eltern zwischen beiden Formen wählen. Der Besuch eines weiterführenden Bildungsganges setzt Eignung voraus.

(2) Die Eignung einer Schülerin oder eines Schülers für einen weiterführenden Bildungsgang ist gegeben, wenn bisherige Lernentwicklung, Leistungsstand und Arbeitshaltung eine erfolgreiche Teilnahme am Unterricht des gewählten Bildungsgangs erwarten lassen.

(3) Bei der Wahl des weiterführenden Bildungsganges haben die Eltern Anspruch auf eingehende Beratung. Sie teilen ihre Entscheidung der Klassenlehrerin oder dem Klassenlehrer der abgebenden Jahrgangsstufe mit. Erfolgt die Wahl des weiterführenden Bildungsganges durch die Wahl der Realschule oder des Gymnasiums oder der entsprechenden Zweige der schulformbezogenen (kooperativen) Gesamtschule, so nimmt die Klassenkonferenz unter dem Vorsitz der Schulleiterin oder des Schulleiters dazu schriftlich Stellung. Die Stellungsnahme muss eine Empfehlung für den Bildungsgang oder die Bildungsgänge enthalten, für den oder für die die Eignung der Schülerin oder des Schülers nach Maßgabe des Abs. 2 gegeben ist. Wird dabei dem Wunsch der Eltern widersprochen, so ist ihnen eine erneute Beratung anzubieten. Halten die Eltern ihre Entscheidung aufrecht, so erfolgt die Aufnahme in den gewählten Bildungsgang.

(4) An schulformabhängigen (integrierten) Gesamtschulen (§ 27) sind die Informations- und Entscheidungsrechte der Eltern bei der Ersteinstufung von Schülerinnen und Schülern in Fachleistungskurse den Vorschriften des Abs. 3 entsprechend zu wahren.

(5) Für die endgültige Entscheidung über den weiteren Bildungsweg am Ende der Förderstufe gilt Abs. 3 Satz 1 bis 5 entsprechend. Der Übergang in den Bildungsgang der Realschule oder des Gymnasiums setzt voraus, dass ihn die Klassenkonferenz der abgebenden Förderstufe befürwortet.

(1) Parents decide on the track choice after elementary school. If the school track is offered in a specific school or within a comprehensive school, parents may choose between these two school types. Ability is required for attending a secondary school track.

(2) Ability is indicated by performance, proficiencies and attitudes anticipating the successful completion of the chosen school track.

(3) Parents have the right to get advice on the school track choice. They inform the previous class teacher about their decision. If the intermediate or higher level secondary school or a corresponding track in a cooperative comprehensive school is chosen, the class conference guided by the school principal needs to provide an item of written comment. This needs to include a recommendation for the track or tracks according to the pupil's abilities as stated in section 2. In case of disagreement with the parents' choice, further advice must be offered to the parents. If the parents adhere to their decision, the pupil is assigned to the respective track chosen by the parents.

(4) Integrated comprehensive schools (§ 27) must consider the parental rights according to section 3 when grouping in ability groups takes place for the first time.

(5) The final decision on tracking after the 'support stage' follows the guidelines of section 3. Transition to the higher level secondary school requires the approval of the 'support stage' class conference.

Chapter 2

Instrumental Variable Estimates of Educational Effects of Age at School Entry in Germany *

Joint work by
Patrick A. Puhani and Andrea M. Weber

* This chapter is based on the discussion paper versions: IZA Discussion Paper, 1827, 2005 and Darmstadt Discussion Paper in Economics, 151, 2005.

Abstract: We estimate the effect of age at school entry on educational attainment using three different data sets for Germany, sampling pupils at the end of primary school, in the middle of secondary school and several years after secondary school. Results are obtained based on instrumental variable estimation exploiting the exogenous variation in month of birth. We find robust and significant positive effects on educational attainment for pupils who enter school at seven instead of six years of age: Test scores at the end of primary school increase by about 0.42 standard deviations and years of secondary schooling increase by almost half a year.

JEL classification: I21, I28, J24

Keywords: education, immigration, policy, identification

Acknowledgement: This project has been initiated through discussions with Michael Fertig, RWI, Essen. We are also grateful to Gianno De Fraja, Peter Fredriksson, Edwin Leuven, Stephen Machin, Kjell Salvanes as well as participants of the IZA Summer School 2005 in *Buch am Ammersee* and the CEPR-IFAU-Uppsala Universitet Second Network Workshop 'Economics of Education and Education Policy in Europe' in Uppsala for helpful comments.
We thank Hans-Peter Hafner of the Research Data Center (*Forschungsdatenzentrum*) of the Statistical Office of the state of *Hessen* for help with the administrative data for *Hessen*.
'The Youth and Young Adult Longitudinal Survey used in this paper has been obtained from the Central Archive for Empirical Social Research (*Zentralarchiv für Empirische Sozialforschung*), University Cologne. The data of the "Youth and Young Adult Longitudinal Survey" has been collected by A. Fischer (Psydata Institute), the *"Jugendwerk der Deutschen Shell"*, R. K. Silbereisen (*Friedrich-Schiller-Universität Jena*), L. A. Vaskovics (*Otto-Friedrich-Universität Bamberg*) and J. Zinnecker (*Universität Siegen*) and prepared and documented by the *Zentralarchiv für Empirische Sozialforschung* (*ZA*). The mentioned persons and institutions are not responsible for the analysis and interpretation of the data in the present study.' All remaining errors are our own.

Table of Contents for Chapter 2

2.1 Introduction to Chapter 2

The ideal age at which children should start school and the effectiveness of pre-school learning programs are subjects of ongoing debates among researchers and policy makers. For example, in the economic literature Currie (2001) summarises evidence on early childhood education. Age at school entry effects are estimated in Angrist and Krueger (1992) and Mayer and Knutson (1999) for the United States, Leuven *et al.* (2004) for the Netherlands, Strøm (2004) for Norway, Bedard and Dhuey (2005) for a set of industrialised countries, Fertig and Kluve (2005) for Germany and Fredriksson and Öckert (2005) for Sweden. In Germany, as in most other European countries, children are traditionally supposed to start school when they are about six years old. A look back in history reveals that starting education at the ages six or seven is not just a feature of the industrialised time. Already in Germany's mediaeval predecessor, the Holy Roman Empire, the track to knighthood began at age seven as a footboy (*Page*).[21] In post-war Germany, the changing attitude towards school entry age has been driven by debates among educationalists. In the beginning of the 1950s, Kern (1951) hypothesised that a higher school entry age could prevent children from failing in school. Subsequently, the school entry age was increased by a total of five months in 1955 and in 1964. Since that time, there has also been a trend to have children with learning problems enter school one year later than recommended by the official school entry rule. In recent years, however, debates on the long duration of the German education system have taken early school entry back on the agenda. Policy makers in Germany's decentralised education system have subsequently implemented measures to reduce the average age at school entry.[22] Therefore it seems reasonable to ask whether such policies can be expected to improve educational attainment.

In this paper, we estimate the causal effect of varying the age at school entry in Germany between six and seven years by an instrumental variable strategy using the exogenous variation of month of birth as an instrument for the age at school entry. The variation between ages six and seven is both a major variation observed internationally for the school starting age and a major issue of discussion in the national German debates. Using three different data sets, we measure the effect of age at school entry at three different stages:

[21] The period as a footboy was followed by the stages varlet (*Knappe*) at age 14 and knight (*Ritter*) at age 21. In contrast to modern times, it was not deemed important for a footboy to know how to read and write.

[22] In many schools, pupils may now enter school when they are five years old. For example, starting in the school year of 2005/2006, Berlin is changing the school entry regulation (*Schulgesetz*), so that the compulsory school entry age is half a year earlier than it used to be. At the same time the possibility to start school at a later than at the prescribed age has been abolished in Berlin.

at the end of primary school, in the middle of the secondary school track and several years after the end of secondary schooling. Our outcome measures are a test score for primary school pupils and the school track attended or accomplished in the latter two data sets, respectively. To the best of our knowledge, ours is the second study investigating the effect of age at school entry by instrumental variable estimation for Germany. Because the previous study by Fertig and Kluve (2005) uses data sampled on the outcome variable (potentially leading to biases), we use two further data sets, one of which contains *all* pupils in a German state.

The influence of school entry age on educational outcomes is a well-discussed topic, especially in the U.S. and British empirical educationalist literature.[23] However, these studies do not sufficiently account for the endogeneity of the age at school entry: In Germany, as well as in many other countries, school entry age is not only determined by some exogenous rule, but depends on the child's intellectual or physical development or the parents' will, too. In several countries (*e.g.* the U.S.) some schools even use standardised tests in order to assess potential first graders' or kindergartners' school readiness.

A key institutional difference between Germany on the one hand and the U.S. or the U.K. on the other is that in Germany each child independently of date of birth has to complete at least nine years of compulsory full-time schooling[24]. In the U.S. and the U.K., length of mandatory schooling varies with date of birth, as children are allowed to leave school once they have reached a certain age (*cf.* Angrist and Krueger, 1992, for the U.S. and Del Bono and Galindo-Rueda, 2004, for the U.K.).[25] Hence, in these Anglo-Saxon countries compulsory

[23] Stipek (2002) provides a thorough review of this literature. One type of existing studies considers the effects of academic red-shirting (*i.e.* the delay of school entry) and early grade retention (*e.g.* May *et al.*, 1995; Jimerson *et al.*, 1997; Zill *et al.*, 1997; Graue and DiPerna, 2000) or of early school admission of selected children (*cf.* Proctor *et al.*, 1986, for a review). However, these studies do not appropriately take the endogeneity problem in measuring entry age effects into account and the mixed findings are therefore hard to interpret (*cf.* Stipek, 2002; Angrist, 2004). A second stream of literature examines the effect of entry age induced through season of birth on educational and social outcomes or mental development (*e.g.* Kinard and Reinherz, 1986; Morrison *et al.*, 1997; Hutchison and Sharp, 1999; Stipek and Byler, 2001). The results mostly indicate that there are no long-lasting effects while there is evidence of positive effects of a higher school entry age in the short run. Since outcomes are separately analysed by season of birth, which is taken as exogenous, the applied methods solve the endogeneity problem by producing reduced form estimates (without however explicitly discussing it). None of the mentioned studies uses an IV approach as in the recent economic literature.

[24] The exact rule depends on the state. The nine or ten years of compulsory full-time education are followed by either at least one additional year of full-time education or by several years of part-time education in a vocational school (*Berufsschule*) within the German apprenticeship system.

[25] To be more precise, in England and Wales children could traditionally (between 1962-1997) leave school at the beginning of the Easter holiday in the school year in which they attained the relevant leaving age if they were born between September and the end of January. Children born between February and the end of August could not leave before the end of May.

schooling length is shorter for pupils having entered school at an older age. In Germany, however, all pupils at least have to wait until their ninth school year has finished before they may leave full-time education. Consequently, the German institutional setup allows identification of age at school entry effects independently of compulsory schooling, which is not possible in the U.S. or the U.K.

A further feature that makes the German case interesting to examine is that the German education system is highly selective. Unlike in most other countries, the child's performance in primary school is crucial for the educational career of a person because at the end of primary school (at age ten; primary school usually lasts for four years) children are selected into one of three educational tracks: the most academic is *Gynmasium*, usually consisting of nine further years of schooling, followed by *Realschule* (six years) and *Hauptschule* (five years and the most vocational track). As track selection is supposed to be based on the pupil's primary school performance, the German track system may aggravate age at school entry effects by perpetuating inequalities arising at early stages of the education system (*cf.* Hanushek and Wößmann, 2005). Hence, age at school entry may have larger and more lasting effects in Germany than in countries with a comprehensive school system.

The paper is structured as follows. Section 2.2 outlines age at school entry regulations for the cohorts we observe in our data and sketches main features of the German school system. The three data sets we use are described in Section 2.3. First, for primary school test scores we rely on the 'Progress in International Reading Literacy Study' of 2001 (PIRLS). Second, for the school track during secondary schooling we use newly available administrative data for the state of *Hessen* including *all* pupils in general education in the school year 2004/2005. Third, for schooling accomplishment several years after compulsory schooling we draw on data from the German Youth and Young Adult Longitudinal Survey of 1991. Section 2.4 argues that our empirical approach to identify the effect of age at school entry on educational outcomes is justified. We show that the instruments are effectively uncorrelated with the observed variables used as regressors and that first-stage regressions do not exhibit a weak instrument problem. The estimation results are presented and discussed in Section 2.5. We find robust evidence that increasing the age at school entry from six to seven years raises primary school test scores by more than two fifths of a standard deviation and increases the amount of secondary schooling by almost half a year (about five months). Only results based on the German Youth and Young Adult Longitudinal Survey are not robust. However, we place less weight on results from this survey as it is sampled on the basis of our outcome variable, which is likely to lead to biased estimates. Section 2.6 concludes and reports results from a small-size survey of headmasters and headmistresses, which we carried out in order to discuss potential explanations for our empirical estimates.

2.2 Age at School Entry and the German Education System

In international comparison, the German compulsory school starting age of six years is equal to the median and mode of the distribution displayed in Table 2.1. Before the age of six, German children usually attend kindergarten, which is a playgroup rather than a pre-school. Projects where children learn how to read and write in kindergarten are recent and rare. Therefore, entering primary school for a German child traditionally has meant moving from a playgroup to an educational regime of teaching from eight o'clock in the morning to 12 o'clock in the afternoon with only short breaks (there is some variation on these times by state).

Although the exact school entry age is regulated by law in Germany, personal and school discretion is high. The school laws (*Schulgesetze*) of the states (*Länder*) are traditionally based on the so-called Hamburg Accord (*Hamburger Abkommen*) which was in place in Western Germany between 1964 and 1997. The Hamburg Accord states that children whose sixth birthday is before the end of June of a given calendar year enter school at the beginning of the corresponding school year (normally in August). Children born later are supposed to start school in the following calendar year (again around August). Thus, the *theoretical school entry age* $I^1(b_i, s_i)$ (as recommended by the Hamburg Accord) is related to a child's month of birth b_i and the month the school year starts s_i in the following way:

$$I^1(b_i, s_i) = \begin{cases} \dfrac{(72 + s_i) - b_i}{12} & if \;\; 1 \le b_i \le 6 \\ \dfrac{(84 + s_i) - b_i}{12} & if \;\; 6 < b_i \le 12 \end{cases}, \qquad (4)$$

where the theoretical school entry age $I^1(b_i, s_i)$ is measured in years (in decimals up to the month). The indicator for the month of birth b_i ranges from one to twelve, whereas the variation in s_i is between seven and nine, as school always started in July, August or September for the cohorts considered in our samples. If b_i and s_i are exogenous (*cf.* Sections 2.3 and 4), the theoretical school entry age $I^1(b_i, s_i)$ is exogenous and can be used as an instrument for the actual age at school entry. Note that the start of the school year s_i varies over calendar year and state as shown in Table 2A.1 in the Appendix (whereas August 1st is the official nationwide school starting date, the actual starting dates vary by calendar year and state in order to avoid traffic jams on the motorways during vacation times).

Contrary to the Hamburg Accord, children born between the official cut-off date 'end of June' and the school year starting date s_i are often admitted to school in the calendar year when they turn six years of age. This practice provides an alternative instrument, which is exogenous under the same conditions as $I^1(b_i, s_i)$:

$$I^2(b_i, s_i) = \begin{cases} \frac{(72 + s_i) - b_i}{12} & if\ 1 \leq b_i \leq s_i \\ \frac{(84 + s_i) - b_i}{12} & if\ \mathrm{s_i} < b_i \leq 12 \end{cases}. \quad (5)$$

As – despite of the Hamburg Accord – the decisions on when to admit children to school are *de facto* taken at the school or parent level, we consider alternative instruments based on other cut-off dates in order to see which instrument works best in the first-stage regression. In addition to the end of June, we define instruments with the end of July until the end of September as cut-off dates c:[26]

$$I^3(b_i, c, s_i) = \begin{cases} \frac{(72 + s_i) - b_i}{12} & if\ 1 \leq b_i \leq c \\ \frac{(84 + s_i) - b_i}{12} & if\ \mathrm{c} < b_i \leq 12 \end{cases}. \quad (6)$$

Although the variation in the three instruments just introduced is mainly driven by variation in month of birth b_i rather than the school starting month s_i, the latter may be a source of endogeneity rendering the instruments invalid. One potential cause for the endogeneity of s_i is that s_i depends on the calendar year and state and both these factors may be correlated with unobserved characteristics relevant to educational outcomes. In addition to that, we observe pupils several years after the start of primary school, such that a pupil may have entered primary school in a different state than the state where he or she is observed in the data. In this case, we do not know the exact school starting date, which may lead to endogeneity of s_i by a correlation of the state we wrongly allocate a mover to and unobserved characteristics of that person.

To avoid these potential endogeneities, we build the following alternative instrument, which does not depend on s_i:

$$I^4(b_i,c)=\begin{cases}\dfrac{(72+8)-b_i}{12} & if\ 1\le b_i\le c\\ \dfrac{(84+8)-b_i}{12} & if\ c<b_i\le 12\end{cases}, \qquad (7)$$

where we substitute '8' for the school starting date s_i, as August is the official start of the school year in Germany around which actual school starting dates vary.

The Hamburg Accord was made less binding in 1997, when the Council of the Ministers of Education encouraged the states to deviate from the traditional school entry cut-off date of end of June and to allow later cut-off dates (up to the end of September). This increased even further the discretion that schools and parents already had *de facto*. For example, in the state of *Hessen*, for which we use recent administrative data, the current official school entry age policy is to generally recommend application of the Hamburg Accord but additionally allow for early entry of children born several months later. In how far actual school starting ages comply with the regulations outlined here will be exhibited in Section 2.4, when we discuss the validity of the instruments introduced above.

Apart from the school entry regulations, tracking is another feature of the German education system important to the analyses in this paper. After four years in primary school, pupils usually change to one of three secondary school tracks.[27] The most vocational and least academic level of secondary schooling is called *Hauptschule* (grades five to nine), the intermediate level *Realschule* (grades five to ten) and the most academic level *Gymnasium* (grammar school, grades five to thirteen).[28] Track selection is important, as only graduation from *Gymnasium* directly qualifies for university or polytechnic tertiary education. *Hauptschule* and *Realschule* are supposed to be followed by vocational training within the German apprenticeship system. The distribution of pupils across the three tracks varies by state, but for Germany as a whole it is about equal. Although there are ways to enter the *Gymnasium* track after *Hauptschule*, *Realschule* or apprenticeship training, the track selection

---

[26] We also tried later cut-off dates up to December, but these did not explain school entry age behaviour well.

[27] In the East German States of *Berlin* and *Brandenburg*, primary school goes up to grade six, so that the selection into school tracks starts two years later there than in the rest of Germany.

[28] In the East German states of *Sachsen* and *Thüringen, Gymnasium* ends after grade twelve. In the small West German state of *Rheinland-Pfalz*, *Gymnasium* nowadays ends after twelve and a half years of schooling. Most states are currently planning to have *Gymnasium* end after grade twelve, but this is not relevant for our samples.

after primary school is a key decision for the economic and social life of a person in Germany (Dustmann, 2004). Note that Germany also has comprehensive schools (*Gesamtschulen*) as well as schools for children with special needs, mostly due to physical or mental disabilities (*Sonderschulen*). There are also so-called Waldorf schools that follow a special pedagogy which does not give marks to pupils, for example. In the year 2003, only 17 percent of graduates came from schools outside of the standard tracking system (eleven percent were in comprehensive schools, six percent in special schools and one percent in Waldorf schools), as Figure 2.1 shows.

2.3 Data

We use three different data sets measuring educational outcomes at three stages of pupils' or young adults' lives. First, the 'Progress in International Reading Literacy Study' (PIRLS) of 2001 provides us with internationally standardised test scores and other relevant information for 6,591 German pupils in the fourth grade of primary school. Second, we use administrative data on all pupils from the state of *Hessen* in the school year 2004/2005 who entered primary school between 1997 and 1999 and were attending secondary school at the time of observation. The observed cohorts overlap with those tested in the PIRLS study.[29] Our estimation sample thus contains 182,676 observations. Finally, the Youth and Young Adult Longitudinal Survey of 1991 provides data on secondary educational achievement for a sample of 1,199 persons aged between 22 and 29 at the date of interview. The surveyed cohorts are thus between about 20 and 30 years older than the persons sampled in the first two data sets. More detail is given in the following sub-sections.

2.3.1 The Progress in International Reading Literacy Study (PIRLS)

The PIRLS data has been collected by the International Association for the Evaluation of Educational Achievement (IEA) and includes test scores of an internationally conducted standardised reading literacy test as well as background information on pupils and parents. For Germany, 7,633 pupils at the end of fourth grade in 211 primary schools are sampled.[30]

[29] We also tried to obtain administrative pupil statistics from other German states, but were either denied access or told that an essential variable for our analysis is missing.

[30] Because the sampling units are schools rather than pupils, all of our results presented in the following sections use standard errors adjusted for clustering. We also use the sampling weights provided in the data set.

Because we lack information on the age at school entry (to the month) for more than one thousand observations, our effective sample size is reduced to 6,591.[31]

As we are interested in estimating the effect of age at school entry on educational attainment, we might like to sample a birth or school entry cohort and estimate the effect of interest after four years of schooling, no matter which grade pupils have achieved by then. The other possibility is to measure educational attainment at the end of primary school irrespective of how long it took the pupil to reach grade four. The advantage of the latter approach is that the pupil's performance at grade four of primary school is what matters in the end for the secondary school track recommendation he or she receives. As the PIRLS data samples pupils in grade four, we can only identify the parameter associated with the latter approach, except that it is not an entry cohort, but an exit cohort (fourth graders at the end of primary school) that is sampled. In our data, 86 percent of pupils have entered school in 1997, whereas eleven and two percent have entered in 1996 (grade repeaters) and 1998 (grade skippers), respectively. Hence, we observe pupils once they have reached grade four, even if they have spent only three or even five years in school. If grade repetition and skipping behaviour has not changed significantly between these neighbouring cohorts, our results should be roughly representative for the 1997 school entrants.

The instruments we can build with the PIRLS data are limited to the $I^4(b_i, c)$-type, because the data does not contain information on the state a pupil lives in (*cf.* Section 2.2). Therefore, the types of instruments using the school year starting month s_i, which depends on the state, cannot be constructed with the PIRLS data.

2.3.2 Administrative Data on All Pupils in the State of *Hessen*

The second data source we use is 'Pupil-Level Data of the Statistics of General Schools for the State of *Hessen*' (*Hessische Schülereinzeldaten der Statistik an allgemein bildenden Schulen*). It covers *all* pupils in general education in the school year 2004/2005 and is collected on behalf of the state Ministry of Education. To our knowledge, this is the first research paper using this individual-level administrative data.

[31] The age at school entry is unfortunately not missing at random: immigrants and pupils whose parents have a comparatively low level of education are overrepresented among the missing observations. If age at school entry is also missing systematically for pupils with unobserved characteristics that are relevant to educational attainment, our estimates based on the selected sample might be biased. However, as we control for parental background and immigrant status, which is likely to be correlated with these characteristics, we hope to reduce this potential bias markedly.

The original data set contains 694,523 observations from 1,869 schools. As it does not contain any school marks or test scores, we use the track attended in 2004/2005 by pupils having entered school between 1997 and 1999 as the outcome variable. This leaves us with 182,676 observations, 93 percent of them in grades six to eight. Tracks are coded according to the years of schooling they imply: 13 for *Gymnasium* (grammar school), 10 for *Realschule* and 9 for *Hauptschule*. Pupils at comprehensive schools (*Gesamtschule*) are frequently allocated to an internal track that corresponds to *Gymnasium*, *Realschule* or *Hauptschule*, as well. In this case, the administrative data codes them as if they were in these schools. If no such information is given, we code them as 10, *i.e.* equivalent to *Realschule*. Pupils in special schools (*Sonderschule*) are allocated code 7.[32]

Although the administrative data for the state of *Hessen* allows the construction of all four types of instruments introduced in Section 2.2, *i.e.* $I^1(b_i,s_i)$, $I^2(b_i,s_i)$, $I^3(b_i,c,s_i)$ and $I^4(b_i,c)$, we have a preference for $I^4(b_i,c)$-type instruments. The reason is that we do not know whether pupils entered schools in the state of *Hessen*, so that the exact school starting month s_i has to be proxied by assuming that pupils entered primary school in the same state where they attend secondary school. However, if we make that assumption, $I^1(b_i,s_i)$ and $I^4(b_i,c)$ as well as $I^3(b_i,c,s_i)$ and $I^4(b_i,c)$ will accidentally be identical, as August was the theoretical school starting date s_i in *Hessen* for the cohorts considered here. If both the cut-off and the school-starting dates are set to August ($c = s_i = 8$), then $I^2(b_i,s_i)$, $I^3(b_i,c,s_i)$ and $I^4(b_i,c)$ are identical.

2.3.3 The Youth and Young Adult Longitudinal Survey

The final data set used in this study is the German Youth and Young Adult Longitudinal Survey *(Jugend- und Junge-Erwachsenen-Längsschnitt).* This data is an extension of the so-called Y*outh 92* (*Jugend 92*) survey conducted by the German Shell Company (*Deutsche Shell AG*). We use the 1991 cross section of this survey because it is – to our knowledge – the only

[32] 0.86 percent of pupils in the original sample are still in primary school when we observe them: they are excluded from the sample in the reported estimates since we do not know which track they will be assigned to. To check in how far these pupils affect our results, we carry out a rather extreme robustness check by allocating code 4 to individuals still in primary school, which indicates the fact that they failed to move to secondary school in time. We carry out a further sensitivity check by excluding pupils in comprehensive and special schools. Pupils in Waldorf schools are not separately identified: they are like comprehensive schools. Note that private schools are included in our sample: 10,709 pupils are in private schools, about 76 percent of whom attend grammar school (*Gymnasium*).

German data set informing on school entry age and educational attainment later in life. However, we also have doubts about the adequacy of this survey for the analysis of age at school entry effects on educational outcomes. The problem is the stratified sampling by gender, region, town size, birth year *and* achieved secondary school track. Stratification on the outcome variable 'school track' is likely to lead to biases. We could not obtain more information on the stratification procedure and therefore have to carry out the analysis with this *caveat* in mind.

Since we want to consider individuals with completed secondary education, we restrict the sample to persons who are between 22 and 29 years old at the time of interview (29 is the maximum age in the survey; our sample thus covers birth cohorts 1961 to 1969). We only consider West Germans, as the school system in Eastern Germany was quite distinct from the one in the West. This leaves us with 1,199 observations. As the data set does not contain weights, we prefer regression specifications where we control for gender, region, town size and birth year, which are all variables influencing the sampling design.

The coding of educational attainment is similar to the one of the administrative data for the state of *Hessen*, *i.e.* 13 years of schooling for *Gymnasium* (grammar school), 10 for *Realschule* and 9 for *Hauptschule*. Persons without any of these degrees are coded as 7.[33]

As to the construction of the instrumental variables, we do not know whether a person went to primary school in a different state from the one where he or she was interviewed in 1991. Therefore, the instruments $I^1(b_i,s_i)$, $I^2(b_i,s_i)$ and $I^3(b_i,c,s_i)$ might be endogenous through a correlation of unobserved skills with state of residence in the survey year 1991, which would translate into a correlation of unobserved skills with the assumed theoretical school starting date s_i and hence the instruments. Note, however, that the variation in s_i is small (only three months) compared to the variation in month of birth b_i (twelve months), which is the main factor driving the variation in the instrument. Therefore, potential biases might be small. Nevertheless, we mainly rely on $I^4(b_i,c)$-type instruments in order to avoid these potential problems.

In the following section, we provide more detail on theoretical and actual age at school entry in our data and further discuss the validity of the instruments.

---

[33] As comprehensive schools (*Gesamtschulen*) were mostly introduced in the 1980s, the birth cohorts 1961 to 1969 did not attend them.

2.4 The Exogeneity of Month of Birth and First-Stage Regressions

2.4.1 The Endogeneity of Age at School Entry

Regressing educational outcomes on age at school entry by ordinary least squares regression (OLS) must be expected to yield biased estimates rather than the causal effect of age at school entry on educational results. The reason is that the school entry decision is influenced not just by regulations like the Hamburg Accord, but also by the child's development as well as the parents' and the school's judgements. Thus, ambitious parents may want to push for an early school entry (at age 5) of their child or children with learning problems might be recommended to enter school one year later (at age 7) than prescribed by official regulations. These mechanisms suggest that on average, less able pupils will enter school at a later age and thus OLS estimates of age at school entry effects on educational attainment should exhibit a downward bias.

Figure 2.2 displays the distributions of the actually observed school entry age and the theoretical entry age according to the 'Hamburg Accord' (for PIRLS, we exhibit the instrument $I^4(b_i, c = 6)$ with the end of June as cut-off date; for the administrative data for *Hessen* as well as the Youth and Young Adult Longitudinal Survey, we display the school entry age according to the Hamburg Accord $I^1(b_i, s_i)$, where we calculate s_i on the assumption that a pupil has not changed state. The prescribed school entry age varies between 6.17 and 7.08 years in the first two data sets (PILRS and administrative data for *Hessen*), whereas it varies between 6.17 and 7.22 years in the Youth and Young Adult Longitudinal Survey. The larger variation in the latter data set is explained by the variation of school starting months s_i across German states.

It is clearly visible that the actual distribution of age at school entry is far more dispersed and skewed to the right than the distribution prescribed by the Hamburg Accord (the skewness is positive and ranges from 0.33 to 0.50 in the three graphs). This is because many parents/schools have children start school one year later than suggested by the regulations. However, a few children also start school one year earlier at about age five. Despite of that, the large majority of pupils start school at the prescribed age.

A further graphical illustration of the degree of compliance with the age at school entry rules discussed in Section 2.2 is provided in Figure 2.3. The first panel displays the actual age at school entry by month of birth together with three different instruments. The instruments are a $I^4(b_i,c)$-type version of the 'Hamburg Accord' using the end of June as cut-off date without knowing the actual school entry month, $I^4(b_i,c=6)$, and two further versions of $I^4(b_i,c)$, one with the end of July and one with the end of August as cut-off date c. Visual inspection suggests a significant correlation between the instruments and the actual age at school entry. However, children born from October to June enter school a little older on average than prescribed by the Hamburg Accord. This is consistent with the graphs in Figure 2.2 showing that late entry is more frequent than early entry. However, for those born between July and September, the average age at school entry is lower than prescribed by the Hamburg Accord illustrating the fact that close to the cut-off point, many parents decide for their children to enter school early. This suggests using instruments with later cut-off dates.

A similar picture concerning non-compliance with the cut-off date of the Hamburg Accord arises in the last two panels of Figure 2.3. In the administrative data for *Hessen*, pupils born just after the cut-off date 'end of June' enter school earlier on average than demanded by the Hamburg Accord. Therefore we also consider instruments of type $I^3(b_i,c,s_i=8)=I^4(b_i,c)$ using the end of July and the end of August as cut-off dates, as in the PIRLS data.[34] Note that as school always started in August for the cohorts we analyse with the administrative data from *Hessen*, $I^2(b_i,s_i=8)=I^3(b_i,c=8,s_i=8)=I^4(b_i,c=8)$ if August is chosen as the cut-off date.

In the Youth and Young Adult Longitudinal Survey (third panel of Figure 2.3), the compliance with the Hamburg Accord, $I^1(b_i,s_i)$, seems weakest of all analysed data. The instruments using the start of the school year in the respective state, $I^2(b_i,s_i)$, as well as $I^4(b_i,c)$ with August or September ($c=8; c=9$) as cut-off dates appear to describe school entry behaviour better.

2.4.2 The Exogeneity of the Instruments

In order for $I^1(b_i,s_i)$ to $I^4(b_i,c)$ to be valid instruments, they have to be both correlated with the actual age at school entry and uncorrelated with unobserved factors influencing educational performance in a prospective regression equation. In order to gauge whether the

[34] In the following analyses using the administrative data for *Hessen* we only report results based on instruments with June and July as cut-off dates. We also tried instruments based on later cut-off dates but there is no relevant compliance with these cut-off dates.

instruments are truly exogenous variables, *i.e.* uncorrelated with any *unobserved* factors that might influence educational performance, an assumption we cannot test directly, we test whether the instruments are correlated with *observed* variables that we believe might influence educational performance. In addition, we rely on 'discontinuity samples' where the included observations are limited to pupils born in the two adjacent months around the cut-off dates.

Table 2.2 lists the groups of regressors that we include in our two-stage least squares (2SLS) instrumental variable estimation models. Note that the regressors enter both in the first-stage (as discussed below in this section) and in the second-stage regressions (as discussed in Section 2.5). The set of variables is partly determined by data availability in the respective data sets. In the first set of regressions ('exo1') we include no regressors in the model except age at school entry as the variable to be instrumented. The justification for this procedure is that if the instrument (mainly driven by month of birth, *cf.* Section 2.2) is completely random and therefore exogenous, no other control variables are required in order to estimate the causal effect of age at school entry on educational attainment consistently in a 2SLS estimation procedure. Nevertheless, control variables that influence educational attainment may reduce the standard errors of the estimates.

As a first extension of the set of regressors ('exo2'), we therefore include gender and regional indicators (if available). In the administrative data for *Hessen*, we also control for the school entry cohort among 'exo2'. In the Youth and Young Adult Longitudinal Survey, year of birth is included among this set of regressors, as it is also one of the variables on which the sample is stratified. The third set of regressors ('exo3') adds cultural background to the set of regressors, measured either by an immigration or nationality indicator, as in PILRS and the administrative data for *Hessen*, or by religion, as in the Youth and Young Adult Longitudinal Survey, which does not contain information on country of origin. The fourth extension ('exo4') adds parental education, which is available in the PIRLS data and the Youth and Young Adult Longitudinal Survey, but not in the administrative data for *Hessen*. The fifth addition ('exo5') is only feasible for the Youth and Young Adult Longitudinal Survey and consists of variables referring to the first school day, *e.g.* an indicator whether the child received a gift from the parents, which is a tradition in Germany. Finally, we add family background variables like books at home or number of siblings to obtain the last set of regressors ('exo6'). This is only possible for the PIRLS data and the Youth and Young Adult Longitudinal Survey, because the administrative data for *Hessen* does not contain this information. We consider the control variables added in 'exo5' and 'exo6' as potentially problematic, as they might be an outcome of pupils' (potential) performance and hence be endogenous: For example, parents might be more likely to give presents to children or to buy them books if they are not (expected to be) performing well in school. Hence, controlling for these sets of variables may take out some of the effect that age at school entry has on educational attainment.

Although low correlations between the instrument and observable variables are supportive of the instrument's exogeneity, they do not provide a guarantee. Therefore, we additionally consider so-called 'discontinuity samples', where only students born in two adjacent months around the respective school entry cut-off points are included. The reason for examining these discontinuity samples are potential direct effects of month or season of birth on health and educational outcomes, as briefly surveyed in Bound, Jaeger and Baker (1995). By restricting our samples to persons with two adjacent months of birth, we hope to eliminate any potential seasonal effects which might affect the validity of the instruments. Furthermore, any differences in parental attitudes potentially reflected in planned timing of births should be minimised for children born in two adjacent months, as it is hard to assure for a child to be born in a very specific month.

In Table 2.3 to Table 2.5 we display the simple correlations between a selected set of instruments and the full set of our control variables for all three data sets. Correlations significant at the 10 or 5 percent level are marked with one or two asterisks, respectively. As Table 2.3 shows, the maximum correlation for the full-sample of the PIRLS data equals 0.02 in absolute value, which is very small. Hence, the few correlations of instruments with regressors that are significantly different from zero are very close to zero. This finding is even more striking in the full sample of the large administrative data set for *Hessen* in Table 2.4: No correlation is larger than 0.01 in absolute value. Our instruments (mainly driven by month of birth) thus seem unrelated to gender, the district of residence and the country of origin. Table 2.3 also shows that our instruments are virtually unrelated to parental education, the number of siblings and the number of books in the household. In the discontinuity samples, there are two statistically significant correlations of 0.05 in the PIRLS data (*cf.* Table 2.3). In the administrative data for *Hessen* (*cf.* Table 2.4), which is larger in sample size, the maximum correlation remains 0.01 in the discontinuity samples. We interpret the findings of no or extremely small correlations of the instruments with the observed characteristics as indications that the instruments are plausibly exogenous and thus not correlated with unobserved variables either.

A *caveat* applies in this respect when considering the findings based on the Youth and Young Adult Longitudinal Survey in Table 2.5. Although by far the largest part of the correlations of the instruments with the regressors is insignificant and small in absolute size, some correlations are as large as 0.06 to 0.08 in the full sample and up to 0.14 in the (small) discontinuity samples in absolute value, respectively. This may be explained by the stratified sampling procedure. As Table 2.5 demonstrates, parental education is one of the variables slightly correlated with the instruments (the correlation is 0.05 in absolute value in the full sample). This is potentially reflecting the fact that sampling is based on the outcome variable 'school track'. Because educational levels of parents and children are known to be correlated (Dustmann, 2004), sampling on the outcome variable must be expected to generate biases. Therefore, we believe that instrumental variable estimates based on the Youth and Young Adult Longitudinal Survey should be regarded with caution. Nevertheless, we will still report results based on this survey, also to compare our findings with those of Fertig and Kluve (2005), who use this data.

2.4.3 First-Stage Regressions

Having discussed the exogeneity of our instruments, we now check the second condition for a valid instrument, namely the (partial) correlation with the variable to be instrumented (age at school entry). Table 2.6 to Table 2.8 report coefficients of the instruments together with the *F*-statistics of the tests for significance of the instruments in the first-stage regressions of the 2SLS estimation procedure. A rule of thumb states that an *F*-statistic below about 10 is indicative of a weak instrument problem (Staiger and Stock, 1997; Stock, Wright and Yogo, 2002).[35] The tables therefore display the *F*-statistics for various specifications distinguished by both the choice of instrument and the choice of regressors ('exo1' to 'exo6') as outlined in Section 2.4.2.

Table 2.6 to Table 2.8 clearly show that, in all three data sets, we have instruments with *F*-statistics largely above the threshold value of 10. The degree of compliance with the

[35] If instruments are weak, the 2SLS estimator has a high standard error and inference using asymptotic approximations for the standard errors is not reliable. Furthermore, already a very small correlation between the instrument and the error term of the outcome equation may lead to significant inconsistencies if instruments are weak (Bound, Jaeger, and Baker 1995). In other words, 2SLS with weak instruments is generally not appropriate.

rules built into the various instruments can be seen from the coefficients reported in the tables. Note that using the full samples, the degree of compliance is influenced by the behaviour of both individuals born around the cut-off date and persons born in months like January, which are quite distant from the alleged cut-off dates. In the PIRLS data (Table 2.6), the coefficients of the full sample vary between 0.31 and 0.49, which means that increasing the prescribed age at school entry by one year raises the actual age at school entry by between a third and a half of a year on average. The 'Hamburg Accord' shows the highest compliance in the PIRLS data.

Using the discontinuity samples of persons born in the two months adjacent to the respective cut-off date also reveals that the compliance with the Hamburg Accord is strongest with a coefficient of 0.40. The cut-off date July renders a weak instrument (with an *F*-statistic of around 5). The compliance is stronger again when August is used as cut-off, with a first-stage coefficient of 0.27 and an *F*-statistic of around 70. Although this makes both the Hamburg Accord (June as cut-off date) and August as cut-off date the relevant instruments (later cut-offs do not render useful instruments), one has to keep in mind that 2SLS estimation identifies the causal effect of age at school entry for 'compliers', *i.e.* those persons who react to variations in the instrument (Imbens and Angrist, 1994). Although the 2SLS model implicitly assumes that the effect of age at school entry is homogeneous across the population, the estimate is an equivalent of the local average treatment effect (LATE) as introduced in Imbens and Angrist (1994) for binary instruments. Persons reacting to June (the Hamburg Accord) as cut-off might consequently be more representative for the average pupil, unlike those reacting to August as the cut-off. It is plausible that the group of pupils born in August and entering school at the age of just about six (younger than prescribed by the Hamburg Accord) are above-average achievers and hence distinct from the representative pupil. If virtually all 'compliers' born in August and September are high achievers, it may be that the 'compliers' for the instrument $I^4(b_i, c)$ with August as the cut-off date are affected differently by the variation in the age at school entry than compliers with the official rule of the Hamburg Accord.[36]

In the administrative data for *Hessen* the degree of compliance is also half a year for the Hamburg Accord in the full sample, but a third of a year for July as the cut-off date (*cf.* Table 2.7). An investigation of the discontinuity samples around the cut-off points reveals that compliance is only sufficiently strong with the Hamburg Accord (June as cut-off) with 0.41 as

[36] We do not know who is a complier: the counterfactual had a pupil been born in a different month is not observable. Thus we cannot test whether the compliers born in August or September are high achievers.

the first-stage coefficient. The first-stage coefficient for July as the cut-off date is very low at 0.04 (so is the coefficient for August as the cut-off date, which is not shown here). The full-sample first-stage results with the July cut-off date thus seem to stem from a broader concept of compliance than the specific cut-off date 'end of July'. This is to say they are explained by the fact that months of birth in spring (*e.g.* May) lead to lower ages of school entry than months of birth in autumn (*e.g.* October), which is true no matter whether June or July are chosen as hypothetical cut-off date. Hence, what we observe in the full-sample with July as the cut-off date is in fact a reflection of the Hamburg Accord (the June cut-off). Therefore, in the administrative data for *Hessen* just as in the PIRLS data, the Hamburg Accord is the appropriate instrument.

In the Youth and Young Adult Longitudinal Survey, the degree of compliance is highest if the school starting months or September are used as cut-off dates (the coefficients vary between 0.74 and 0.79 in these cases, meaning that an increase in the prescribed school entry age by one year increases the average age at school entry by about three quarters of a year). The coefficient in the full sample reduces to 0.58 or 0.59 if August is used as the cut-off month. However, the discontinuity samples reveal compliance only for September as the cut-off date. The displayed coefficients vary between 0.66 and 0.72. First-stage coefficients of other cut-off dates are not significant and not shown here.

In sum, the estimated first-stage coefficients and their *F*-statistics confirm the picture given in Figure 2.2 that compliance with the school entry rules is considerable, but not perfect. As mentioned above in this section, 2SLS estimates the effect of age at school entry on educational attainment only for the group of persons complying with the regulations. Therefore, we have to keep in mind that the results discussed in the following section may not be representative for the pupil population as a whole (*cf.* Imbens and Angrist, 1994, on local average treatment effects). Non-compliers are likely to be particularly weak pupils who enter school later than prescribed or strong performers who enter school earlier than suggested by the rules, or children of parents who have strong views on the age at which their child should enter school and consequently would not respond to cut-off dates.

Having justified our instruments in terms of exogeneity and (partial) correlation with the age at school entry, we present the results of the second stage of the 2SLS estimates in the following section.

2.5 The Effect of Age at School Entry on Educational Outcomes

2.5.1 Ordinary Least Squares Results

Table 2.9 to Table 2.11 report the estimated effects of age at school entry on educational attainment from regressions with different sets of control variables ('exo1' in the first line indicating no control variables, and the last line indicating the full set of control variables as listed in Table 2.2 to Table 2.5).

The columns headed '(0)' of Table 2.9 to Table 2.11 report OLS regression coefficients for the full samples. In all data sets, the regression coefficient is negative and significantly different from zero if no control variables are included (specifications 'exo1'). This means that educational attainment and age at school entry are negatively correlated: Pupils who enter school at a later age achieve less than their peers entering at a younger age. However, as we include more and more control variables into the regressions, the OLS-coefficients decrease in absolute value in all data sets indicating that actual age at school entry is influenced by factors relevant to educational performance. This is highly suggestive of age at school entry being an endogenous variable, which warrants instrumental variable estimation.

2.5.2 Two-Stage Least Squares Results

What happens to the estimated effect of age at school entry on educational attainment if we apply 2SLS estimation with the instruments tested in Section 2.4? A glance at Table 2.9 to Table 2.11 reveals first that instrumental variable estimation switches the sign of the estimated effect from negative to positive in all data sets in all specifications with statistically significant coefficients. Second, in the PILRS data and in the administrative data for *Hessen*, the 2SLS estimates with the Hamburg Accord, the strongest instrument, are all positive and significantly different from zero. Third, the differences of the point estimates between the full and the discontinuity samples are smaller than a discontinuity-sample standard deviation in both the PIRLS and administrative data for *Hessen* if the Hamburg Accord (June as the cut-off date) is chosen as instrument (the specification with control variables 'exo2' in the data for *Hessen* is the only exception where this difference is slightly larger). Fourth, although the size of the estimated effects varies by the choice of instrument (as can be expected from the first-stage results discussed in Section 2.4.3), they hardly vary by the choice of control variables (*i.e.* between specifications 'exo1' to 'exo6'/'exo3') in Table 2.9 and Table 2.10: Indeed, the variation of the 2SLS estimates within a column is virtually always less than any estimated standard error of a coefficient in that column. In the Youth and Young Adult Longitudinal Survey (Table 2.11) this is only true in column (4a), where $I^4(b_i,c)$ is used as instrument with September as cut-off date, the appropriate cut-off date for this data.

In the following, we will discuss the 2SLS results in detail by data set. The PIRLS data do not contain information on the state a pupil lives in, so that we can only use instruments of type $I^4(b_i,c)$ for this data. Column (1a) in Table 2.9 sets $c = 6$ in the full sample, which is the same cut-off date as in the Hamburg Accord. The results for the corresponding discontinuity sample are shown in column (1b). Columns (2a), (3a) and (2b), (3b) shift the cut-off date to July and August for the full and the discontinuity samples, respectively. As reasoned in Section 2.4.2, the inclusion of more control variables in the 2SLS regressions mostly reduces the standard error of the estimated coefficient on age at school entry (as we move from 'exo1' to 'exo6').

The main finding in Table 2.9 is that the estimated effect of age at school entry on educational outcomes varies from 25.8 to 30.7 test scores in columns (1a) and (1b) when June is used as the cut-off date. The ranges of the estimated effects overlap between the full sample in column (1a), 26.8 to 30.7 test scores, and the discontinuity sample in column (1b), 25.8 to 29.0 test scores. The estimates for the strongest instrument are therefore robust across the full and the discontinuity samples.

In the discontinuity samples, we find no significant effect when July or August are used as cut-off dates. Given the weak instrument for July as cut-off, this is not surprising. However, the instrument with August as the cut-off is not weak, which has been shown in Table 2.6. As argued in Section 2.4.3, the 'compliers' reacting to June and August as cut-off dates may be rather different groups, with the latter plausibly comprising more talented pupils (the latter group consists of pupils who enter school early even if they are born two months after the official cut-off date). Because 2SLS estimates a local average treatment effect, the effect for the group of compliers with respect to the August cut-off may be different from the effect for the group of compliers with respect to the Hamburg Accord (June cut-off). Given the results based on the discontinuity samples, the statistically significant full-sample results in columns (2a) and (3a) in Table 2.9 are likely to stem from compliance of individuals born in months distant from the respective cut-off dates and hence seem to be driven by the cut-off date June, *i.e.* the Hamburg Accord (*cf.* our discussion in Section 2.4.3). We thus regard the results based on the Hamburg Accord as our most relevant estimates. Nevertheless, the findings on the compliers reacting to the August cut-off indicate that age at school entry effects are heterogeneous across the pupil population.

How can the results be interpreted? A representative estimate from the Hamburg Accord as instrument is an increase in test scores of around 27 points for raising the school entry age by one year (from about six to seven years of age). This is a bit more than two fifths of the standard deviation of test scores in PIRLS (the standard deviation is 63.61, so that the estimated effect amounts to 0.42 standard deviations). More intuition for the size of this effect is derived from a comparison of the differences in test scores between the different German school tracks in the PISA 2000 study (where ninth graders' reading literacy is tested).[37] In the PISA data for ninth graders, the differences in test scores are 0.78 standard deviations between pupils in *Gymnasium* and *Realschule* and 1.01 standard deviations between *Realschule* and *Hauptschule* (Baumert *et al.*, 2003). Therefore, our estimates imply that increasing the age at school entry from six to seven years increases reading literacy by more than half of the difference between the average *Gymnasium* track and the average *Realschule* track performance. This is quite a substantial effect and indicates that age at school entry may influence track choice, as also shown in the following paragraphs.

Table 2.10 presents the effects of age at school entry on track attendance in the middle of secondary school (measured by the number of school years associated with each track as outlined in Section 2.3.2). Results are based on administrative data for the state of *Hessen.* The 2SLS estimates with the Hamburg Accord $I^1(b_i,s_i)$ as instrument are given in columns (1a) and (1b) for the full and the discontinuity samples, respectively. The results for $I^4(b_i,c)$-type instruments with July as the cut-off dates are shown in columns (2a) and (2b), respectively. As discussed in Section 2.4.3, there is hardly any compliance with the July cut-off in this data, so that columns (2a) and (2b) are only displayed for illustrative purposes. We do not consider them indicative of the causal effect of school entry age on track attendance, since they are based on weak instruments.

Because the administrative data for *Hessen* is large in terms of number of observations (in fact we observe the population), the reported 'standard errors' in Table 2.10 all indicate significance (only the standard error in column (2b) is sizeable because compliance with the corresponding instrument is very low, *cf.* Table 2.7). As to the estimated effect of age at

[37] We do not use the PISA data for our estimations, because it does not contain the required information.

school entry on educational attainment using the Hamburg Accord as instrument, columns (1a) and (1b) yield comparable estimates in the ranges of 0.41 to 0.45 and 0.37 to 0.40 for the full and discontinuity samples, respectively, with minor variation among specifications with different sets of control variables.[38] The Hamburg Accord as the appropriate instrument thus exhibits robust positive effects of age at school entry on track attainment across full and discontinuity samples as well as across specifications with different sets of control variables: Entering school at the age of seven rather than six raises secondary schooling by almost half a year (around five months).[39] This effect is implied if a deferral of school entry by one year increases the probability of attending *Gymnasium* instead of *Realschule* by about 13 percentage points.[40] Thus the results from the largest data set used in this study are qualitatively consistent with the findings from primary school reading literacy test scores.

Table 2.11 presents the estimated coefficients based on the third and smallest data set, the Youth and Young Adult Longitudinal Survey. As we have outlined above, the Youth and Young Adult Longitudinal Survey is sampled on the outcome variable (educational achievement), which may generate biases. This fact or simply the lower sample size in this data, and hence the relatively large standard errors, may explain the comparatively wide variation of the estimated effects of age at school entry on educational attainment across different specifications concerning the set of control variables.[41] Due to these data deficiencies, the results from this data set can only be taken with a grain of salt.

In contrast to the findings from the first two data sets, expanding the set of regressors in the Youth and Young Adult Longitudinal Survey changes (mostly reduces) the coefficient

[38] The reported coefficients would be similar but somewhat higher if we did not exclude persons still in primary school from the sample. If we include primary school pupils (with code 4 as the outcome, *cf.* footnote 32), the coefficients related to columns (1a) and (1b) range between 0.46 to 0.49 and 0.43 to 0.46 in the full and discontinuity samples, respectively. Hence, early school entry seems to increase the likelihood of repeating grades in primary school. As a further robustness check we exclude pupils in comprehensive and special schools (*Gesamtschule* and *Sonderschule*). In this case the effects are only slightly different from the presented effects and range between 0.42 and 0.47 (1a) and 0.36 to 0.39 (1b) in the full and discontinuity samples, respectively.

[39] This interpretation implies the assumption that pupils will complete the track which they attend in the middle of secondary school, when we observe them.

[40] The estimated effect is potentially driven by both increases in the probability to attend *Realschule* rather than *Hauptschule* and increases in the probability to attend *Gymnasium* rather than *Realschule*. In order to find out which of these effects drives the results, we estimated linear probability models of *Gymnasium* versus *Realschule/Hauptschule* attendance as well as of *Gymnasium/Realschule* versus *Hauptschule* attendance. Estimates were obtained by two-stage least squares (2SLS) using the same instruments and control variables as in Table 2.10. The results show robust increases of *Gymnasium* versus *Realschule/Hauptschule* attendance by between 11 and 13 percentage points and increases of *Gymnasium/Realschule* versus *Hauptschule* attendance of about 2 to 3 percentage points. The numbers are very robust and significant across different specifications (using the Hamburg Accord as instrument). Hence, it seems that the age at school entry matters for achieving *Gymnasium* attendance, which is the step towards university education and high labour market returns.

of age at school entry on educational attainment and all effects are insignificant in the specifications with the full set of regressors ('exo6'). It is interesting, though, that the 2SLS point estimates based on the full sample without control variables ('exo1') are in a similar range, *viz.* from 0.37 to 0.77, to the results based on the administrative data for *Hessen*, albeit with much larger standard errors. Column (1a) displays the results with the Hamburg Accord as the instrument, $I^1(b_i,s_i)$, column (2a) with the school starting dates, $I^2(b_i,s_i)$, and columns (3a) and (4a) show the estimates based on instrument $I^4(b_i,c)$ with August and September as cut-off dates, respectively. Estimation results for the discontinuity sample with September as the cut-off date are shown in column (4b). As we do not know whether a person has changed the state of residence since entering primary school (an event 15 to 23 years before the survey date), the instruments $I^1(b_i,s_i)$ and $I^2(b_i,s_i)$ might be affected by measurement error as discussed in Section 2.3.3. Thus focusing on the $I^4(b_i,c)$-type instruments, we observe a remarkable fall in the point estimates in columns (3a), (4a) and also (4b) when including parental education among the set of regressors, *i.e.* when moving from 'exo3' to 'exo4'. Considering the findings with 'exo4' as the control variables (*i.e.* without first school day and family background variables, which might be endogenous as argued in Section 2.4.2), we find an estimated effect of 0.47 years in column (3a), which is significant at the 10 percent level. This estimate is in line with the results from the administrative data from *Hessen* and provides very tentative evidence that age at school entry might have long-run effects on educational achievement.

2.5.3 Results for Subgroups

Having established robust evidence from PIRLS and the administrative data from *Hessen* that an older age at school entry raises educational attainment, we carry out a subgroup analysis in Table 2.12 and Table 2.13 for these two data sets. Due to the smaller sample size and the reservations we have concerning the quality of the Youth and Young Adult Longitudinal Survey, we do not consider this data set for a subgroup analysis.

Table 2.12 displays first-stage coefficients and *F*-Statistics as well as second-stage estimation results for native males, native females, immigrant males, immigrant females and for pupils with parents with and without an academic degree, respectively. The estimates are

[41] For the discontinuity sample, the point estimates are negative, but insignificant due to large standard errors.

exhibited both for the full and the discontinuity samples with the Hamburg Accord as the instrument and refer to the specification with all control variables (‘exo6’).

The main results from the subgroup analysis based on the PIRLS data are that German males benefit more than German females from later school entry: Coefficients are 42.9 (standard error 8.6) versus 16.2 (standard error 8.4) in the full samples, respectively. Due to smaller sample sizes and large standard errors (the latter ranging from 5.9 to 62.1 test scores), the subgroup estimates, especially in the discontinuity samples, are generally harder to pin down. Potentially for the same reasons, some estimated effects for male immigrants (full sample), female immigrants (full and discontinuity sample), for female natives (discontinuity sample) and for pupils with parents holding an academic degree (discontinuity sample) are not significantly different from zero.

Note that only the effects for the group of persons who comply with the instruments in the respective subgroup are identified by 2SLS. Therefore, the estimated ‘local average treatment effects’ do not have to be representative for the subgroups in general (for example, if most immigrant males enter school at the age of seven anyway, the compliers will be a small and unrepresentative group). However, first-stage coefficients show that the degree of compliance is similar for most subgroups, especially in the full sample. First-stage coefficients in the full sample mainly range between 0.44 and 0.56. Exceptions are immigrant females and pupils whose parents have attained an academic degree, for whom compliance is somewhat lower (the full-sample first-stage coefficients for these two groups are 0.38 and 0.35, respectively).

As in Table 2.12 for the PIRLS data, the estimates in Table 2.13 are shown both for the full and for the discontinuity samples with the Hamburg Accord as instrument and refer to the specification with all control variables (‘exo3’ in this case). The subgroup results for the administrative data for the state of *Hessen* do not confirm that German males benefit more from later school entry than German females. However, the different results from these two data sets need not contradict as PIRLS measures only reading literacy, whereas the secondary school track in the data for *Hessen* is a more general indicator for educational attainment.

In the administrative data for *Hessen*, we can distinguish between different groups of nationalities (German, Turkish, predominantly Muslim countries without Turkey, Italy/Greece and former Yugoslavia). As sample sizes for all subgroups except Germans and Turks are below 1,600 (full samples) or 300 (discontinuity samples), the standard errors of the second-stage estimates range between 0.26 and 0.76, so that second-stage coefficients for these nationality groups are hard to pin down. We therefore ignored other nationality groups with even smaller sample sizes.

The first-stage coefficients for almost all subgroups are close to those of the sample as a whole, exceptions being both males and females from predominantly Muslim countries without Turkey, where compliance is lower (full-sample first-stage coefficients range between 0.35 and 0.36 for these groups compared to between 0.45 and 0.52 for the rest). Although there is some indication based on the first-stage *F*-statistics that the instruments for these two groups are not that strong, the marginally (in)significant point estimates for females from predominantly Muslim countries without Turkey tentatively suggest that they benefit more than natives from a later age at school entry. However, the large standard errors associated with these estimates make this interpretation somewhat speculative as the difference in the estimated effects is not statistically significant. The somewhat smaller point estimates for Turkish than native pupils are also associated with a sizeable standard error making this difference statistically insignificant. We cannot detect any significant effects of age at school entry for male or female pupils from Italy and Greece or for males from former Yugoslavia. However, at least in the full sample, the estimated effect for females from former Yugoslavia is significant and the largest of all groups (0.95), albeit with a sizeable standard error (0.41).

2.6 Conclusions of Chapter 2

Based on instrumental variable estimation, we recover positive and statistically significant effects on educational outcomes for entering school at a relatively higher age in the current German school system. In the fourth grade of primary school, we find a large effect of about 0.42 standard deviations improvement in the PIRLS test score if the pupil enters at about the age of seven rather than six (*i.e.* a year later according to the school entry rule). This amounts to more than half of the difference in the average *Gymnasium* versus *Realschule* test scores in the OECD PISA study. Administrative data for the state of *Hessen* suggest that the effect of age at school entry persists into secondary school by prolonging average years of schooling by almost half a year (about five months). As the statutory length of *Gymnasium* is mostly 13 years versus ten years in *Realschule*, our estimates of the age at school entry effect are smaller in secondary than in primary school. Yet they remain sizeable.

Compared to Fredriksson and Öckert (2005) and Bedard and Dhuey (2005), who apply an instrumental variable strategy similar to ours to Swedish administrative and the international TIMSS data, respectively, the results for Germany are comparable in size: Fredriksson and Öckert (2005) report that entering school a year later increases ninth graders' grade point average by about 0.2 standard deviations. Similarly, the effects reported in Bedard and Dhuey (2005) range from 0.1 to 0.35 standard deviations for fourth and eighth graders in the countries investigated, most of which are in Europe. Strøm (2004) estimates an effect of 0.17 standard deviations for 15-16 year olds in the Norwegian PISA study, arguing that age at school entry is exogenously driven by regulations in Norway.[42] However, these and our estimates differ from those of Angrist and Krueger (1992) and Mayer and Knutson (1999) for the United States, where either no or negative effects for late school entry are reported. The findings for the U.S. can only be partly explained by the fact that quarter of birth in the U.S., unlike in Germany, affects the duration of compulsory schooling: No and negative effects of later school entry are found for persons having obtained post-compulsory schooling in Angrist and Krueger (1992) and Mayer and Knutson (1999), respectively.

Like Fertig and Kluve (2005), we can hardly detect effects of age at school entry on the educational achievement of young adults based on the German Youth and Young Adult Longitudinal Survey. However, we believe that the sampling procedure in this survey may lead to biased estimates and therefore we place little weight on estimates from this data set.

Given the current trend in Germany to have pupils start school earlier, we interviewed 25 primary school headmasters or headmistresses in the state of *Hessen* by telephone. We asked them about their views on our finding that late school entry improves educational performance.[43] Of the 25 schools, two were operating under a special regime where pupils

[42] Our estimates based on the PIRLS data (0.42 standard deviations) are on the high end of the range of results from other countries. However, in relation to the first-stage coefficients reported for eleven countries in Table 5 of Bedard and Dhuey (2005) as well as those in Fredriksson and Öckert (2005) for Sweden, the degree of compliance with the instrument in Germany is at the very low end in international comparison. As we can only estimate a local average treatment effect, the compliers in Germany might be less representative of the average pupil in Germany than in Sweden, for example, were compliance is higher. This might be one reason – apart from differences in school systems, data collection and other factors – why point estimates differ across countries. Indeed, correlating first- and second-stage coefficients for the eleven countries analysed in Table 5 of Bedard and Dhuey (2005) provides a correlation of -0.20 for science and -0.01 for maths test scores in the TIMSS study. Hence, at least for maths, estimates based on a larger degree of compliance seem to be associated with a lower average treatment effect. The fact that the compliers with the August cut-off in the German PISA data exhibit no significant effect of age at school entry on test scores is consistent with these arguments. We thank Peter Fredriksson for pointing this issue out to us.

[43] We drew 30 telephone numbers of primary schools from the school registry of *Hessen* until we managed to talk to 25 of them (three schools refused to be interviewed by telephone and in two of the schools we could not reach a contact person after several trials).

enter school at the age of five, but with extra logopedic, German language and nursery teacher support. In these schools, five-year olds do not enter grade one, but 'grade zero', which is a mixture between a kindergarten and a school regime. Both schools are satisfied with this regime, as they are able to correct deficits some children have through the extra teaching and nursery resources they have (one of these schools stated that they have a 75 percent immigrant share). In a third school, we were not able to communicate the substance of our question. However, in the remaining 22 'standard' primary schools, 95 percent of headmasters or headmistresses (21 out of 22) said they found our results 'plausible'. We then went on to ask them what they believed could be the reasons for these findings. All 95 percent (21 schools) made statements along the lines that older pupils are more mature, are more able to concentrate when having to keep still in the classroom for long periods of time, are more able to organise themselves (like keeping their belongings together), are less distracted by play and find it easier to overcome frustration. Only 18 percent of schools (four out of 22) felt that relative age effects matter, too. The other schools, however, explicitly denied the importance of relative age effects and stressed that it is personal maturity that matters.[44] Most 'standard' primary schools were opposed to early school entry in the current 'standard' educational regime, but supported the idea of early school entry if the school system changed to a situation similar to the special regime schools, which have extra support for pupils with learning, language or social problems and a 'grade zero' which combines learning with kindergarten elements.

Interpreting the results of our statistical analysis and those of the school survey, we thus find consistent evidence against early school entry into the current German school regime. However, our results should not be interpreted as evidence against early learning. Early learning, organised differently from the standard German school system, might be promising. Which type of early learning works best is an interesting research agenda for the future, once state governments decide to collect and make available appropriate data in this respect. The research presented in this paper suggests that simply reducing the age at school entry in Germany without adapting the style of teaching in early grades is not an optimal strategy. Although it is true that negative effects of early school entry have to be weighed against the economic gains of entering the labour market earlier, new data on earnings, age at

[44] This is consistent with the findings of Fredriksson and Öckert (2005) that absolute age effects dominate relative age effects in Sweden. Since our pupil data do not allow us to distinguish between absolute and relative age effects, we can only present our telephone survey results on this matter.

school entry and educational outcomes for a representative sample would be needed to carry out an appropriate cost-benefit analysis on this issue.

References for Chapter 2

Angrist, J.D. (2004): American Education Research Changes Tack, *Oxford Review of Economic Policy* 20, 198-212.

Angrist, J.D. and A.B. Krueger (1992): The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples, *Journal of the American Statistical Association* 87: 328-335.

Baumert, J., U. Trautwein and C. Artelt (2003): Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens, in Deutsches PISA-Konsortium (eds): *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Verlag Leske + Budrich, Opladen, 261-331.

Bedard, K. and E. Dhuey (2005): *The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects*, Working Paper: Department of Economics University of California, Santa Barbara.

Bertram, T. and C. Pascal (2002): *Early Years Education: An International Perspective*, Qualifications and Curriculum Authority, London.

Bound, J., D.A. Jaeger and R.M. Baker (1995): Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak, *Journal of the American Statistical Association* 90: 443-450.

Currie, J. (2001): Early Childhood Education Programs, *Journal of Economic Perspectives* 15: 213-238.

Del Bono, E. and F. Galindo-Rueda (2004): *Do a Few Months of Compulsory Schooling Matter? The Education and Labour Market Impact of School Leaving Rules*, IZA Discussion Paper No. 1233.

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

Fertig. M and J. Kluve (2005): *The Effect of Age at School Entry on Educational Attainment in Germany*, IZA Discussion Paper No. 1507.

Fredriksson, P. and B. Öckert (2005): *Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labour Market Performance*, IZA Discussion Paper No. 1659.

Graue, M.E. and J. DiPerna (2000): Redshirting and Early Retention: Who Gets the “Gift of Time” and What Are Its Outcomes?, *American Educational Research Journal* 37: 509-534.

Hanushek, E.A. and L. Wößmann (2005): *Does Educational Tracking Affect Performance and inequality?: Differences-in-Differences Evidence across Countries*, NBER working paper series 11124.

Hutchison, D. and C. Sharp (1999): *A Lasting Legacy? The Persistence of Season of Birth Effects*, NFER Conference Paper, presented at the British Educational Research Association Conference, University of Brighton, September.

Imbens, G.W. and J.D. Angrist (1994): Identification and Estimation of Local Average Treatment Effects, *Econometrica* 62: 467-475.

Jimerson, Sh., E. Carlson, M. Rotert, B. Egeland and L.A. Sroufe (1997): A Prospective, Longitudinal Study of the Correlates and Consequences of Early Grade Retention, *Journal of School Psychology* 35: 3-25.

Kern, A. (1951): *Sitzenbleiberelend und Schulreife*, Verlag Herder, Freiburg.

Kinard, E.M. and H. Reinherz (1986): Birthdate Effects on School Performance and Adjustment: A Longitudinal Study, *Journal of Educational Research* 79: 366-372.

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink (2004): *New evidence on the effect of time in school on early achievement*, HEW 0410001, Economics Working Paper Archive at WUSTL.

May, D.C., D.K. Kundert, and D. Brent (1995): Does Delayed School Entry Reduce Later Grade Retentions and Use of Special Education Services?, *Remedial and Special Education* 16: 288-294.

Mayer, S. and D. Knutson (1999): Does the Timing of School Affect How Much Children Learn?, in S. Mayer and P. Peterson (eds): *Earning and Learning: How Schools Matter*, Brookings Institution Press, Washington D.C.

Morrison, F.J., E.M. Griffith and D.M. Alberts (1997): Nature-Nurture in the Classroom: Entrance Age, School Readiness, and Learning in Children, *Development Psychology* 33: 254-262.

Proctor, T.B., K.N. Black and J.F. Feldhusen (1986): Early Admission of Selected Children to Elementary School: A Review of the Research Literature, *Journal of Education Research* 80, 70-76.

Sharp, C. (2002): *School Starting Age: European Policy and Recent Research*, NFER Conference Paper, presented at the LGA seminar 'When Should Our Children Start School?', LGA Conference Centre, London.

Staiger, D. and J.H. Stock (1997): Instrumental Variables Regression with Weak Instruments, *Econometrica* 65: 557-586.

Stipek, D (2002): At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents, *Social Policy Report* 16, 3-16.

Stipek, D. and P. Byler (2001): Academic achievement and social behaviors associated with age of entry into kindergarten, *Applied Developmental Psychology* 22: 175-189.

Stock, J.H., J.H. Wright and M. Yogo (2002): Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics* 4: 518-529.

Strøm, B. (2004): *Student achievement and birthday effects*, Mimeo, Norwegian University of Science and Technology.

Zill, N., L.S. Loomis and J. West (1997): *National Household Education Survey. The Elementary School Performance and Adjustment of Children Who Enter Kindergarten Late or Repeat Kindergarten: Findings from National Surveys*, NCES Statistical Analysis Report 98-097, Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Tables and Figures for Chapter 2

Table 2.1: Compulsory school starting age by country

Age 4	Age 5	Age 6	Age 7
Northern Ireland	Australia (Tasmania)	Austria	Bulgaria
Netherlands (from 8/02)	England	Australia*	Canada
	Malta	Belgium	Denmark
	Netherlands (until 8/02)	Cyprus	Estonia
	New Zealand	Czech Republic	Finland
	Scotland	France	Latvia
	Wales	Germany	Poland
		Greece	Romania
		Hong-Kong	Singapore
		Hungary	Sweden
		Iceland	*Switzerland*
		Republic of Ireland	
		Italy	
		Japan	
		Korea	
		Liechtenstein	
		Lithuania	
		Luxembourg	
		Norway	
		Portugal	
		Slovakia	
		Slovenia	
		Spain	
		Switzerland	
		U.S.A.	

Note: Based on information from 2002. * Except the state of Tasmania. In *Switzerland* entry age differs by region.
Sources: Sharp (2002) and Bertram and Pascal (2002).

Table 2.2: Variables included in the regression models

Group of Regressors	PIRLS 2001	Administrative Data for *Hessen*	Youth and Young Adult Longitudinal Survey
Exo1	Entry age only	Entry age only	Entry age only
Exo2	Exo1 + Gender	Exo1 + Gender + Entry cohorts + County indicators	Exo1 + Sampling variables [a]
Exo3	Exo2 + Cultural variables (immigrant [b])	Exo2 + Cultural variables (country of origin)	Exo2 + Cultural variables (religion [c])
Exo4	Exo3 + Parental education [d]		Exo3 + Parental education [d]
Exo5			Exo4 + First school day variables [e]
Exo6	Exo4 + Family background [f]		Exo5 + Family background [g]

Note: [a] Sampling variables include: dummy variables for gender, year of birth, region, and city size. [b] Immigrant background is controlled for by a dummy variable indicating whether the student or his/her parents were born abroad or if the student often speaks a foreign language at home. [c] We use information on whether individuals characterize themselves as being (1) Christian, (2) believing in a non-Christian religion or (3) not religious. [d] Three categories of parental education are defined: (1) academic education, (2) non-academic degree, (3) no vocational degree. [e] Includes dummy variables on whether the children received a gift at the first school day and whether the parents attended the school entrance ceremony. [f] Includes the number of siblings and its square and the number of books at home. [g] Includes the number of siblings and its square and home/parental background variables (see Table 2.5 for details).

Table 2.3: Simple correlations between instruments and observables (PIRLS)

Sample	Full Sample: born January-December			Discontinuity Samples: born June/July	born July/Aug.	born Aug./Sept.
Instruments / Observable Characteristics	*I4* (cut-off: June)	*I4* (cut-off: July)	*I4* (cut-off: August)	*I4* (cut-off: June)	*I4* (cut-off: July)	*I4* (cut-off: August)
Added in Exo2: Gender (Reference = Female):						
Male	0.00	0.00	0.01	0.03	-0.02	0.04
Added in Exo3: Immigration (Reference = No immigrant background):						
Immigrant	0.00	-0.01	0.00	0.04	-0.02	-0.00
Missing: Immigrant	-0.02	0.01	0.00	-0.03	0.05**	-0.03
Added in Exo4: Parental Education (Reference = No vocational degree)						
Father: Academic degree	0.00	0.00	-0.02	0.00	0.03	-0.03
Mother: Academic degree	0.00	0.01	0.01	-0.02	0.01	-0.00
Father: Non-academic degree	0.01	-0.02	-0.02**	0.03	-0.03	-0.03
Mother: Non-academic degree	0.00	-0.02*	-0.02	0.02	-0.03	0.00
Missing: Education of Father	0.00	0.01	0.02*	-0.03	0.00	0.03
Missing: Education of Mother	0.00	0.01	0.01	-0.01	0.01	-0.01
Added in Exo6: Family Background						
Number of siblings	0.01	0.01	-0.01	-0.01*	0.03	-0.02
Missing: Number of siblings	-0.02**	0.00	-0.01	-0.05	0.05*	-0.01
Log number of books at home	0.01	0.00	-0.02	0.02	0.01	-0.02
Missing: Log number of books	-0.01	0.01	0.02	-0.03	0.02	0.00
Number of observations	*6,591*	*6,591*	*6,591*	*1,123*	*1,226*	*1,262*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2. Exo1 includes only the age at school entry.

Source: PIRLS 2001. Own calculations.

.

Table 2.4: Simple correlations between instruments and observables (Administrative data for *Hessen*)

Sample	Full Sample: born January-December		Discontinuity Samples: born June/July	born July/Aug.
Instruments / Observable Characteristics	*I1* = I4 (Hamb. Accord)	*I3=I4* (cut-off: July)	*I1* = I4 (Hamb. Accord)	*I3=I4* (cut-off: July)
Added in Exo2: Gender (Reference = Female), Entry Cohort (Refer. = 1997) and County Indicators:				
Gender dummy variable (Male = 1)	0.00	0.00*	0.00	0.01
School entry in 1998	0.01**	0.01**	0.00	0.01
School entry in 1999	0.00*	0.00	0.01*	-0.01
County indicator 1 (Darmstadt)	0.00	0.00	0.00	0.00
County indicator 2 (Frankfurt)	0.00	0.00	0.01	0.00
County indicator 3 (Offenbach Stadt)	0.00	0.00*	0.00	0.01
County indicator 4 (Wiesbaden)	-0.01**	-0.01**	0.00	-0.01*
County indicator 5 (Bergstraße / Odenwald)	0.01**	0.00	0.01	-0.01
County indicator 6 (Darmstadt-Dieburg)	0.00	0.00	-0.01	0.00
County indicator 7 (Groß-Gerau)	-0.01**	0.00	-0.01**	0.01**
County indicator 8 (Hochtaunus)	0.00	0.00	0.00	0.00
County indicator 9 (Main-Kinzig)	0.00	0.00	0.00	0.00
County indicator 10 (Offenbach)	0.00**	0.00	0.00	0.00
County indicator 11 (Rheingau-Taunus)	0.00	0.00	0.00	0.00
County indicator 12 (Offenbach)	0.00	0.00	0.00	0.00
County indicator 13 (Wetterau)	0.00	0.00	0.00	0.00
County indicator 14 (Gießen)	0.00	0.00	0.00	0.00
County indicator 15 (Lahn-Dill)	0.00	0.00	0.00	0.00
County indicator 16 (Limburg-Weilburg)	0.00**	0.00	0.01**	0.01*
County indicator 17 (Marburg-Bied./Vogelsb.)	0.00	0.00	-0.01	0.01*
County indicator 18 (Kassel Stadt)	0.00**	0.01**	0.00	0.01
County indicator 19 (Fulda / Hersfeld-Rotenb.)	0.00	0.00	-0.01	0.00
County indicator 20 (Kassel/Werra-Meißner)	0.00	0.00	0.00	0.00
County indicator 21 (Schwalm-Ed./Waldeck-F.)	0.00	0.00	0.00	0.00
Added in Exo3: Country of origin:				
Country 1 (German speaking countries)	0.01**	0.01**	0.00	0.00
Country 2 (Turkey)	-0.01**	-0.01**	0.00	-0.01
Country 3 (Italy and Greece)	-0.01**	0.00	-0.01**	0.01
Country 4 (Former Yugoslavian states)	0.00	0.00	0.01	0.00
Country 5 (Remaining „Western“ countries)	0.00	0.00	-0.01	0.00
Country 6 (Eastern Europe; former Soviet Un.)	0.00	0.00	0.00	0.00
Country 7 (Remaining Muslim countries)	0.00**	0.00	0.00	0.00
Country 8 (Remaining Asia)	0.00	0.00	0.00	0.00
Country 9 (Remaining countries)	0.00	0.00	0.00	0.00
Number of observations	*182,676*	*182,676*	*32,059*	*32,760*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo3) are explained in Table 2.2. Exo1 includes only the age at school entry.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2004/2005 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 2.5: Simple correlations between instruments and observables (Youth and Young Adult Longitudinal Survey Data)

Sample	Full Sample: born January-December			Born Sept./Oct.
Instruments / Observable Characteristics	*I1* (Hamburg Accord)	*I2* (School Year Starting Dates)	*I4* (Cut-off: September)	*I4* (Cut-off: September)
Added in Exo2: Sampling Variables:				
Male (Reference = Female)	0.01	-0.07**	-0.06**	-0.06
Year of birth 1962	0.02	-0.03	-0.03	-0.02
Year of birth 1963	0.02	-0.02	-0.02	-0.09
Year of birth 1964	0.03	-0.01	0.00	0.09
Year of birth 1965	0.06**	0.01	-0.03	-0.09
Year of birth 1966	0.05	0.02	-0.03	-0.09
Year of birth 1967	-0.04	0.02	0.02	0.11
Year of birth 1968	-0.03	0.01	0.03	-0.03
Region 1 (Schleswig-Holstein)	0.03	0.04	0.02	0.03
Region 2 (Bremen)	-0.02	0.03	0.03	0.02
Region 3 (Hamburg)	0.02	0.04	0.02	0.03
Region 4 (Niedersachsen)	-0.04	-0.01	0.02	0.05
Region 5 (Nordrhein-Westfalen)	-0.07**	0.00	0.01	0.04
Region 6 (Hessen)	0.01	0.05	0.04	0.00
Region 7 (Rheinland-Pfalz)	0.02	-0.01	0.01	-0.01
Region 8 (Saarland)	0.02	0.03	0.05*	0.02
Region 9 (Baden-Württemberg)	0.01	-0.03	-0.03	0.02
Region 10 (Bavaria)	0.06**	-0.04	-0.07**	-0.14*
Region 11 (West Berlin)	-0.03	-0.01	0.00	-0.02
< 2,000 residents	0.02	0.04	0.01	-0.05
2,000 - 4,999 residents	-0.01	-0.08**	-0.06**	-0.13*
5,000 - 19,999 residents	0.04	0.00	-0.02	0.01
20000-49,999 residents	-0.02	0.01	0.02	0.11
50,000 - 99,999 residents	0.00	0.00	0.03	0.08
100,000 - 499,999 residents	-0.03	0.00	0.00	0.11
> 499,999 residents	0.00	0.02	0.03	0.07
Added in Exo3: Cultural Variables (Reference = Not religious):				
Christian religion	-0.02	0.00	-0.01	-0.06
Non-Christian religion	0.00	-0.02	0.00	0.00
Added in Exo4: : Parental Education (Reference = No vocational degree):				
Father: Academic degree	0.00	0.04	0.05*	0.09
Mother: Academic degree	-0.05*	0.02	0.04	0.06
Father: Non-academic degree	0.01	-0.01	-0.03	-0.04
Mother: Non-academic degree	0.02	0.01	-0.01	-0.09
Added in Exo5: First School Day Variables:				
Gift received at first school day	0.00	-0.03	-0.01	-0.03
Missing information: Gift received	-0.01	0.04	0.02	0.03
Parents attended entry ceremony	-0.05*	0.01	0.04	0.10
Added in Exo6: Family Background:				
Number of siblings	0.02	-0.05	-0.05*	-0.06
Parental attitude dummy [a)]	0.00	0.03	0.02	-0.04
Home resources indicator [b)]	-0.02	0.08**	0.07**	0.06
Joint activities dummy [c)]	0.00	0.01	-0.02	-0.14*
Number of observations	*1,199*	*1,199*	*1,199*	*173*

Note: [a)] Indicating whether parents had ambitious perceptions concerning child's future when aged 6-12. [b)] Taking the value of one if person read books/magazines at home when aged 6-12. [c)] Indicating whether parents undertook joint leisure time activities (e.g. music, sports) together with child when she was 6-12 years old. * Significant at ten percent level. ** Significant at five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2. Exo1 includes only age at school entry.
Sources: Data of the German Youth and Young Adult Longitudinal Survey, data on school starting dates. Own calculations.

Table 2.6: First-stage results (PIRLS)

Sample	Full Sample: born January-December			Discontinuity Samples: born June/July	born July/August	born August/Sept.
Instruments/ Specifications	(1a) *I4* (“Hamburg Accord”)	(2a) *I4* (cut-off: July)	(3a) *I4* (cut-off: August)	(1b) *I4* (“Hamburg Accord”)	(2b) *I4* (cut-off: July)	(3b) *I4* (cut-off: August)
Exo1	0.49**	0.35**	0.32**	0.40**	-0,10**	0.27**
(*F*-statistic)	(433.1)	(209.3)	(176.4)	(86.7)	(5.3)	(73.3)
Exo2	0.49**	0.35**	0.32**	0.40**	-0,10**	0.27**
(*F*-statistic)	(427.1)	(213.5)	(177.4)	(89.1)	(5.5)	(71.2)
Exo3	0.49**	0.35**	0.32**	0.40**	-0,10**	0.27**
(*F*-statistic)	(426.5)	(216.2)	(175.5)	(90.6)	(5.4)	(70.3)
Exo4	0.49**	0.35**	0.32**	0.40**	-0,10**	0.27**
(*F*-statistic)	(440.8)	(218.9)	(174.0)	(94.6)	(5.8)	(70.4)
Exo6	0.49**	0.35**	0.31**	0.40**	-0.09**	0.26**
(*F*-statistic)	(428.6)	(210.9)	(170.6)	(95.1)	(5.1)	(68.0)
Observations	*6,591*	*6,591*	*6,591*	*1,123*	*1,226*	*1,262*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2.

Source: PIRLS 2001. Own calculations.

Table 2.7: First-stage results (Administrative Data for *Hessen*)

Sample	Full Sample: born January-December		Discontinuity Samples: born June/July	born July/August
Instruments/ Specifications	(1a) *I1*=I4 (Hamburg Accord)	(2a) *I3=I4* (cut-off: July)	(1b) *I1=I4* (Hamburg Accord)	(2b) *I3=I4* (cut-off: July)
Exo1	0.49**	0.32**	0.41**	0.04**
(*F*-statistic)	(8196.0)	(3456.7)	(2277.1)	(21.6)
Exo2	0.49**	0.32**	0.41**	0.04**
(*F*-statistic)	(8189.0)	(3443.0)	(2306.4)	(19.6)
Exo3	0.49**	0.32**	0.41**	0.04**
(*F*-statistic)	(8321.2)	(3499.6)	(2325.5)	(20.3)
Observations	*182,676*	*182,676*	*32,059*	*32,760*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo3) are explained in Table 2.2.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2004/2005 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 2.8: First-stage results (Youth and Young Adult Longitudinal Survey Data)

Sample	Full Sample: born January-December				Discontinuity Sample: born Sept./Oct.
Instruments/ Specifications	(1a) *I1* (Hamburg Accord)	(2a) *I2* (School year starting dates)	(3a) *I4* (cut-off: August)	(4a) *I4* (cut-off: September)	(4b) *I4* (cut-off: September)
Exo1	0.08	0.79**	0.59**	0.75**	0.72**
(*F*-statistic)	(1.8)	(150.0)	(80.1)	(139.2)	(35.7)
Exo2	0.04	0.78**	0.58**	0.74**	0.66**
(*F*-statistic)	(0.6)	(147.5)	(75.5)	(136.8)	(35.8)
Exo3	0.04	0.78**	0.58**	0.74**	0.68**
(*F*-statistic)	(0.6)	(148.2)	(75.3)	(136.9)	(35.6)
Exo4	0.04	0.78**	0.59**	0.74**	0.70**
(*F*-statistic)	(0.6)	(153.5)	(78.3)	(141.6)	(38.1)
Exo5	0.05	0.78**	0.59**	0.74**	0.70**
(*F*-statistic)	(0.7)	(153.1)	(78.1)	(139.7)	(37.8)
Exo6	0.05	0.78**	0.59**	0.74**	0.66**
(*F*-statistic)	(0.8)	(152.3)	(79.3)	(139.0)	(26.8)
Observations	*1,199*	*1,199*	*1,199*	*1,199*	*173*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2.

Sources: Data of the German Youth and Young Adult Longitudinal Survey, data on school starting dates. Own calculations.

Table 2.9: OLS and second-stage results (PIRLS)

Sample	FullSample: born January-December				Discontinuity Samples: born June/July	born July/Aug.	born Aug./Sept.
	(0) OLS	(1a) *I4* ("Hamburg Accord")	(2a) *I4* (cut-off: July)	(3a) *I4* (cut-off: August)	(1b) *I4* ("Hamburg Accord")	(2b) *I4* (cut-off: July)	(3b) *I4* (cut-off: August)
Exo1	-12.80**	30.74**	29.43**	4.34	28.17**	-42.16	-7.53
(s.e.)	(3.0)	(6.2)	(8.5)	(11.5)	(13.2)	(52.5)	(20.9)
Exo2	-11.49**	30.64**	30.20**	6.02	28.18**	-42.15	-5.90
(s.e.)	(3.0)	(6.3)	(8.5)	(11.6)	(13.1)	(52.4)	(21.3)
Exo3	-8.65**	27.14**	27.52**	4.20	28.98**	-46.50	-9.31
(s.e.)	(2.7)	(6.2)	(7.9)	(10.5)	(12.6)	(50.7)	(20.1)
Exo4	-4.57**	27.37**	31.29**	10.12	26.41**	-54.00	-3.10
(s.e.)	(2.3)	(5.8)	(7.8)	(10.2)	(11.5)	(48.8)	(18.8)
Exo6	-1.24	26.77**	30.14**	11.55	25.83**	-43.70	0.34
(s.e.)	(2.2)	(5.6)	(7.2)	(9.9)	(11.2)	(50.2)	(19.5)
Obs.	*6,591*	*6,591*	*6,591*	*6,591*	*1,123*	*1,226*	*1,262*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2.

Source: PIRLS 2001. Own calculations.

Table 2.10: OLS and second-stage results (Administrative Data for *Hessen*)

Sample	Full Sample: born January-December			Discontinuity Sample: born June/July	Discontinuity Sample: born July/August
	(0) OLS	(1a) *I1=I4* (Hamburg Accord)	(2a) *I3=I4* (cut-off: July)	(1b) *I1=I4* (Hamburg Accord)	(2b) *I3=I4* (cut-off: July)
Exo1	-0.37**	0.45**	0.48**	0.40**	1.37**
(s.e.)	(0.01)	(0.03)	(0.04)	(0.05)	(0.66)
Exo2	-0.36**	0.44**	0.49**	0.38**	1.57**
(s.e.)	(0.01)	(0.03)	(0.04)	(0.05)	(0.71)
Exo3	-0.31**	0.41**	0.45**	0.37**	1.49**
(s.e.)	(0.01)	(0.03)	(0.04)	(0.05)	(0.67)
Obs.	*182,676*	*182,676*	*182,676*	*32,059*	*32,760*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo3) are explained in Table 2.2.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2004/2005 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 2.11: OLS and second-stage results (Youth and Young Adult Longitudinal Survey Data)

Sample	Full Sample: born January-December					Discontinuity Sample: born Sept./Oct.
	(0) OLS	(1a) *I1* (Hamburg Accord)	(2a) *I2* (School year starting dates)	(3a) *I4* (cut-off: August	(4a) *I4* (cut-off: September)	(4b) *I4* (cut-off: September)
Exo1	-0.21**	0.47	0.37	0.77**	0.47*	-0.01
(s.e.)	(0.08)	(2.04)	(0.23)	(0.32)	(0.25)	(0.39)
Exo2	-0.21**	0.40	0.27	0.68**	0.46*	-0.16
(s.e.)	(0.09)	(3.33)	(0.23)	(0.32)	(0.24)	(0.44)
Exo3	-0.21**	0.43	0.27	0.68**	0.46*	-0.23
(s.e.)	(0.09)	(3.33)	(0.23)	(0.32)	(0.25)	(0.42)
Exo4	-0.17**	0.88	0.12	0.47*	0.27	-0.46
(s.e.)	(0.08)	(3.22)	(0.21)	(0.28)	(0.22)	(0.39)
Exo5	-0.16**	0.50	0.14	0.50*	0.30	-0.40
(s.e.)	(0.08)	(2.90)	(0.21)	(0.28)	(0.22)	(0.39)
Exo6	-0.14*	0.22	0.09	0.39	0.29	-0.31
(s.e.)	(0.08)	(2.65)	(0.20)	(0.27)	(0.22)	(0.44)
Obs.	*1,199*	*1,199*	*1,199*	*1,199*	*1,199*	*173*

Note: * Significant at the ten percent level. ** Significant at the five percent level. The different specifications (Exo1 – Exo6) are explained in Table 2.2.
Sources: Data of the German Youth and Young Adult Longitudinal Survey, data on school starting dates. Own calculations.

Table 2.12: Subgroup results for the PIRLS data

	First-Stage		Second-Stage	
Male – Native	Full sample	0.45**	Full sample	42.86**
	(*F*)	(138.9)	(s.e.)	(8.6)
(Full sample: 2,642 observations;	born June/July	0.30**	born June/July	59.83**
born June/July: 447 observations)	(*F*)	(21.6)	(s.e.)	(22.5)
Female - Native	Full sample	0.56**	Full sample	16.23**
	(*F*)	(244.7)	(s.e.)	(8.4)
(Full sample: 2,717 observations;	born June/July	0.52**	born June/July	7.25
born June/July: 469 observations)	(*F*)	(104.5)	(s.e.)	(12.8)
Male – Immigrant	Full sample	0.44**	Full sample	20.50
	(*F*)	(33.4)	(s.e.)	(20.2)
(Full sample: 668 observations;	born June/July	0.43**	born June/July	67.38*
born June/July: 109 observations)	(*F*)	(17.7)	(s.e.)	(36.2)
Female - Immigrant	Full sample	0.38**	Full sample	37.65
	(*F*)	(10.8)	(s.e.)	(30.0)
(Full sample: 564 observations;	born June/July	0.30**	born June/July	-4.06
born June/July: 98 observations)	(*F*)	(4.6)	(s.e.)	(62.1)
Parents: Academic Degree	Full sample	0.35**	Full sample	29.36*
	(*F*)	(45.2)	(s.e.)	(17.0)
(Full sample: 1,330 observations;	born June/July	0.29**	born June/July	32.11
born June/July: 223 observations)	(*F*)	(10.1)	(s.e.)	(30.5)
Parents: No Academic Degree	Full sample	0.53**	Full sample	25.71**
	(*F*)	(438.6)	(s.e.)	(5.9)
(Full sample: 5,261 observations;	born June/July	0.43**	born June/July	24.14**
born June/July: 900 observations)	(*F*)	(97.1)	(s.e.)	(11.6)

Note: Effects for the full specifications (Exo6) using the Hamburg Accord based instrument. * Significant at the ten percent level. ** Significant at the five percent level. *F* refers to the *F*-statistics of joint significance of the instruments in the first-stage regressions.
Source: PIRLS 2001. Own calculations.

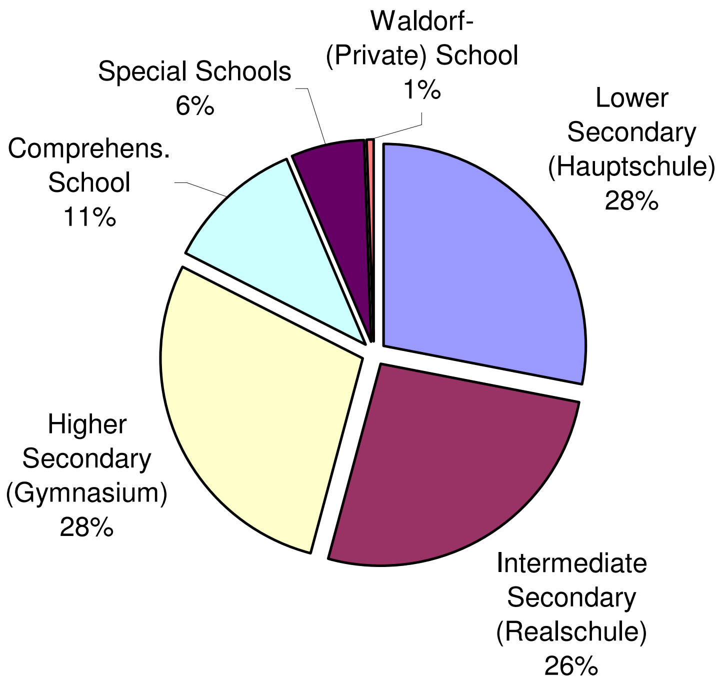
Table 2.13: Subgroup results for the administrative data for Hessen

	First-Stage		Second-Stage	
Male – Native (German speaking	Full sample	0.50**	Full sample	0.41**
countries)	(*F*)	(3885.8)	(s.e.)	(0.04)
(Full sample: 79,400 observations;	born June/July	0.41**	born June/July	0.35**
born June/July: 13,898 observations)	(*F*)	(1025.0)	(s.e.)	(0.08)
Female – Native (German speaking	Full sample	0.50**	Full sample	0.45**
countries)	(*F*)	(3845.2)	(s.e.)	(0.04)
(Full sample: 77,106 observations;	born June/July	0.41**	born June/July	0.39**
born June/July: 13,555 observations)	(F)	(1039.2)	(s.e.)	(0.08)
Male – Turkish	Full sample	0.46**	Full sample	0.21
	(*F*)	(221.0)	(s.e.)	(0.14)
(Full sample: 5,772 observations;	born June/July	0.42**	born June/July	0.33
born June/July: 1,009 observations)	(F)	(62.5)	(s.e.)	(0.23)
Female - Turkish	Full sample	0.49**	Full sample	0.32**
	(*F*)	(255.5)	(s.e.)	(0.13)
(Full sample: 5,647 observations;	born June/July	0.45**	born June/July	0.32
born June/July: 1,045 observations)	(F)	(88.3)	(s.e.)	(0.22)
Male – Predominantly Muslim	Full sample	0.36**	Full sample	0.37
Countries (without Turkey)	(*F*)	(25.0)	(s.e.)	(0.41)
(Full sample: 1,539 observations;	born June/July	0.31**	born June/July	-0.24
born June/July: 247 observations)	(F)	(6.2)	(s.e.)	(0.72)
Female - Predominantly Muslim	Full sample	0.35**	Full sample	0.55
Countries (without Turkey)	(*F*)	(26.3)	(s.e.)	(0.40)
(Full sample: 1,474 observations;	born June/July	0.43**	born June/July	1.00*
born June/July: 248 observations)	(F)	(16.0)	(s.e.)	(0.55)
Male - Italy/Greece	Full sample	0.52**	Full sample	-0.16
	(*F*)	(86.9)	(s.e.)	(0.26)
(Full sample: 1,462 observations;	born June/July	0.37**	born June/July	0.34
born June/July: 271 observations)	(F)	(22.5)	(s.e.)	(0.61)
Female – Italy/Greece	Full sample	0.51**	Full sample	-0.07
	(*F*)	(67.1)	(s.e.)	(0.27)
(Full sample: 1,419 observations;	born June/July	0.50**	born June/July	-0.57
born June/July: 244 observations)	(F)	(31.3)	(s.e.)	(0.44)
Male - Former Yugoslavia	Full sample	0.46**	Full sample	0.04
	(*F*)	(48.9)	(s.e.)	(0.34)
(Full sample: 1,217 observations;	born June/July	0.51**	born June/July	0.01
born June/July: 213 observations)	(F)	(20.1)	(s.e.)	(0.51)
Female - Former Yugoslavia	Full sample	0.45**	Full sample	0.95**
	(*F*)	(46.2)	(s.e.)	(0.41)
(Full sample: 1,190 observations;	born June/July	0.38**	born June/July	1.09
born June/July: 221 observations)	(*F*)	(15.7)	(s.e.)	(0.76)

Note: Effects for the full specifications (Exo3) using the Hamburg Accord based instrument. * Significant at the ten percent level. ** Significant at the five percent level. *F* refers to the *F*-statistics of joint significance of the instruments in the first-stage regressions.

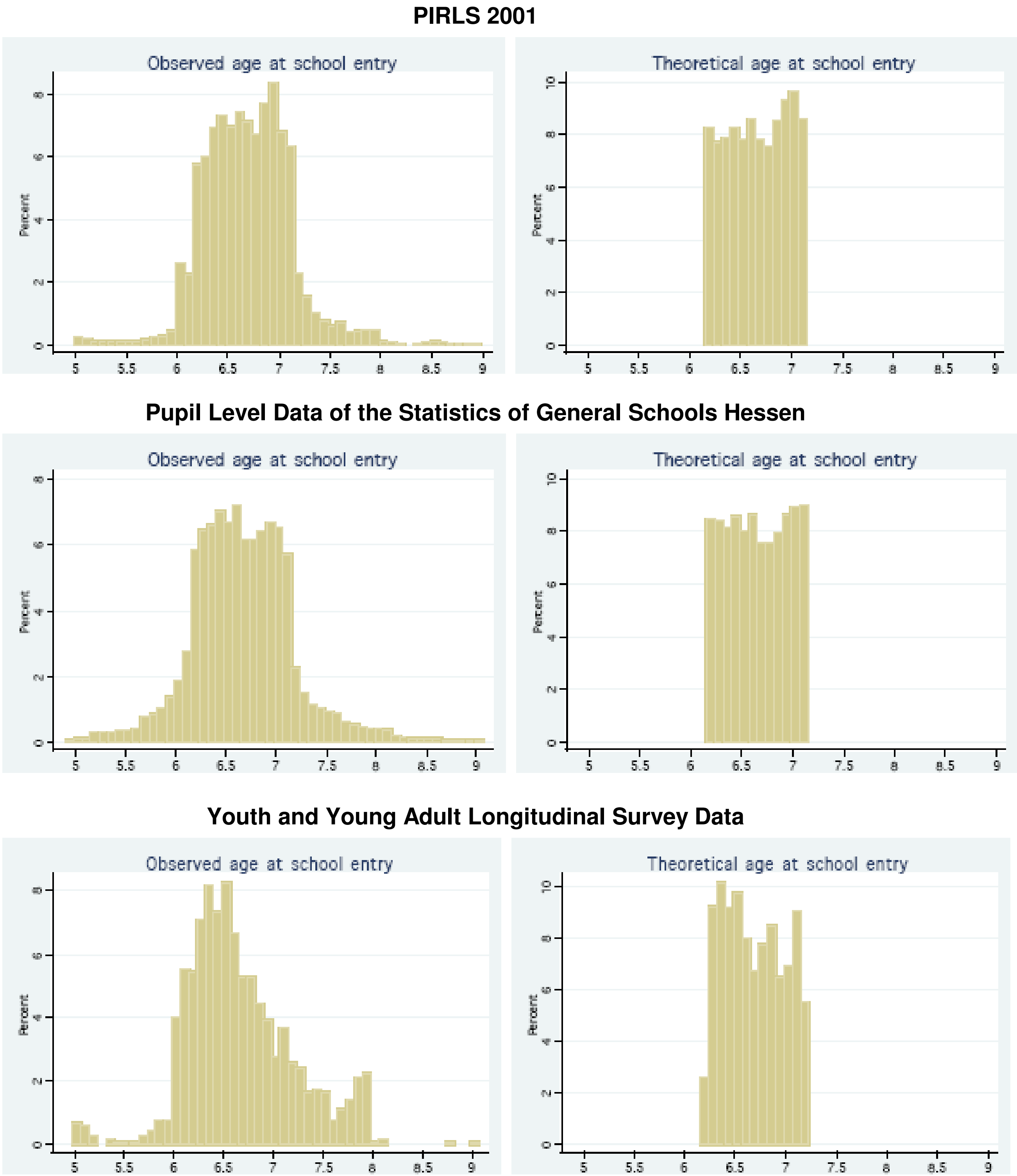
Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2004/2005 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*), data on school starting dates. Own calculations.

Figure 2.1: The German tracking system: Graduates in 2003



Source: German Federal Statistical Office (2004): *Fachserie 11 / Reihe 1: Bildung und Kultur, Schuljahr 2003/04*, Wiesbaden.

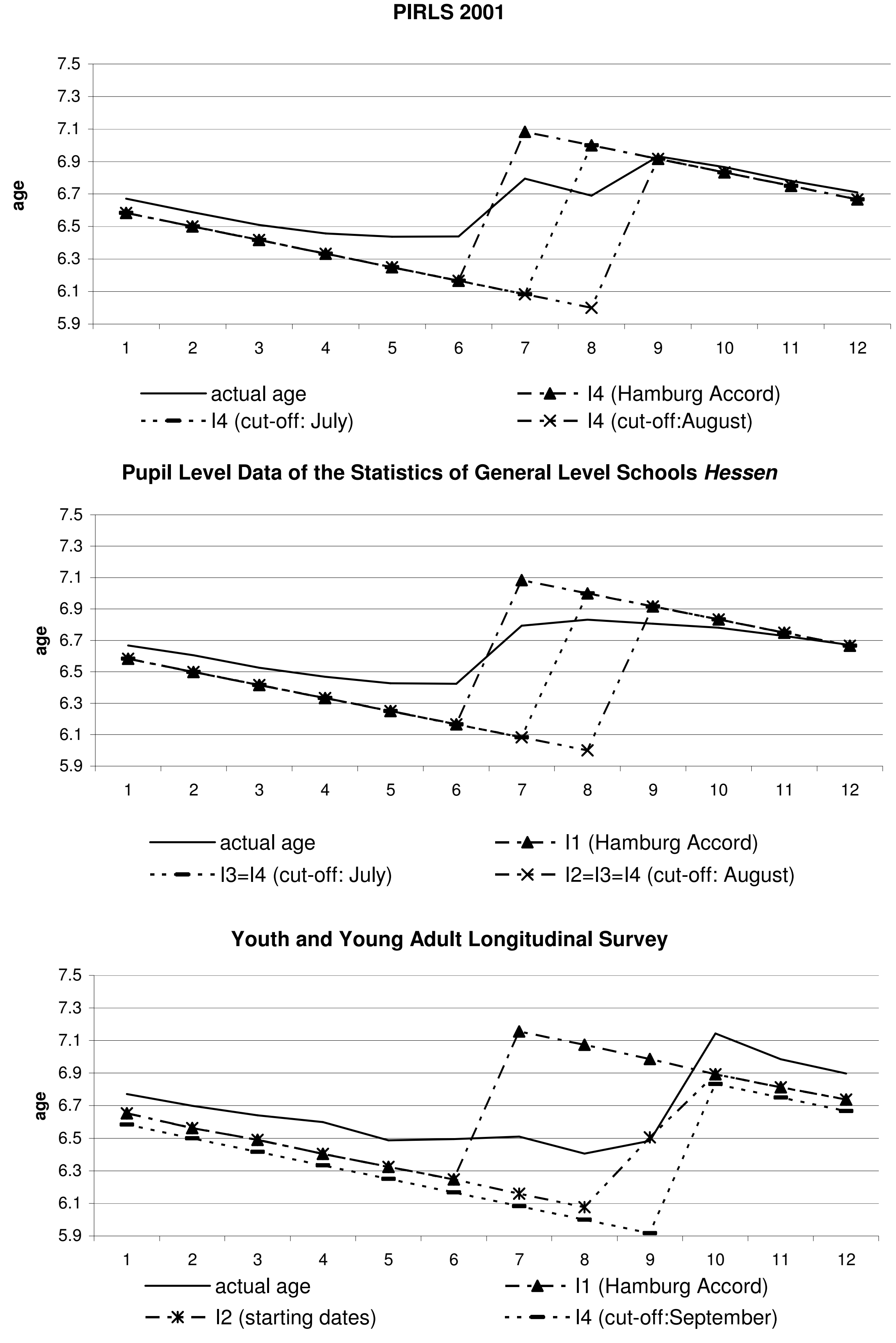
Figure 2.2: Observed and theoretical age at school entry



Note: Theoretical age at school entry according to the 'Hamburg Accord' (June cut-off date).

Sources: PIRLS 2001. Pupil-Level Data of the Statistics of General Schools for the State of Hessen provided by the State Statistical Office (Hessisches Statistisches Landesamt). Youth and Young Adult Longitudinal Survey. Data on school starting dates. Own computations.

Figure 2.3: Observed and theoretical entry ages by birth month



Note: Since there are no state identifiers in the PIRLS data we can only calculate $I^4(b_i, c)$ in this case.

Sources: PIRLS 2001. Pupil-Level Data of the Statistics of General Schools for the State of *Hessen*. Youth and Young Adult Longitudinal Survey. Data on school starting dates. Own computations.

Appendix to Chapter 2

Table 2A.1: Start dates of the school years in the West German states 1966-1976

	Baden-Wurttemberg	Bavaria	Berlin	Bremen	Hamburg	Hessen	Niedersachsen	Nordrhein-Westfalen	Rheinland-Pfalz	Saarland	Schleswig-Holstein
1966	3-Sep-66	5-Sep-66	20-Aug-66	23-Aug-66	15-Aug-66	16-Aug-66	10-Aug-66	7-Sep-66	23-Aug-66	1-Sep-66	9-Aug-66
1967	6-Sep-67	6-Sep-67	31-Aug-67	26-Aug-67	13-Aug-67	31-Aug-67	14-Aug-67	6-Sep-67	22-Aug-67	2-Sep-67	29-Aug-67
1968	7-Sep-68	10-Sep-68	24-Aug-68	22-Aug-68	24-Aug-68	3-Sep-68	26-Aug-68	8-Aug-68	27-Aug-68	31-Aug-68	27-Aug-68
1969	6-Sep-69	10-Sep-69	23-Aug-69	9-Aug-69	9-Aug-69	3-Sep-69	11-Aug-69	3-Aug-69	27-Aug-69	30-Aug-69	9-Aug-69
1970	8-Sep-70	8-Sep-70	22-Aug-70	8-Aug-70	22-Aug-70	27-Aug-70	5-Aug-70	5-Sep-70	26-Aug-70	20-Aug-70	20-Aug-70
1971	13-Sep-71	6-Sep-71	21-Aug-71	4-Sep-71	21-Aug-71	25-Aug-71	8-Sep-71	14-Aug-71	25-Aug-71	1-Sep-71	21-Aug-71
1972	12-Sep-72	18-Sep-72	19-Aug-72	2-Sep-72	26-Aug-72	26-Aug-72	30-Aug-72	5-Aug-72	16-Aug-72	19-Aug-72	26-Aug-72
1973	3-Sep-73	12-Sep-73	25-Aug-73	25-Aug-73	18-Aug-73	18-Aug-73	22-Aug-73	28-Jul-73	8-Aug-73	11-Aug-73	18-Aug-73
1974	24-Aug-74	16-Sep-74	17-Aug-74	17-Aug-74	10-Aug-74	10-Aug-74	14-Aug-74	7-Sep-74	31-Jul-74	3-Aug-74	10-Aug-74
1975	16-Aug-75	10-Sep-75	9-Aug-75	9-Aug-75	2-Aug-75	2-Aug-75	6-Aug-75	30-Aug-75	31-Jul-75	31-Jul-75	2-Aug-75
1976	14-Aug-76	15-Sep-76	7-Aug-76	7-Aug-76	31-Jul-76	28-Jul-76	4-Aug-76	28-Aug-76	8-Sep-76	11-Sep-76	31-Jul-76

Source: Council of the Education Ministers (*Kultusministerkonferenz*).

Chapter 3

How Persistent Is the Age At School Entry Effect in a System of Flexible Tracking? *

Joint work by
Patrick A. Puhani and Andrea M. Weber

* This chapter is based on the following discussion paper versions: IZA Discussion Paper, 2965, 2007; Discussion Paper of the Department of Economics, Leibniz Universität Hannover, 370, 2007; University of St. Gallen, Department of Economics Working Paper Series, 30, 2007.

Abstract: In Germany, students are streamed into secondary school types at the age of ten. This initial tracking decision can be revised at later stages of secondary education. Using administrative data on the population of students in the German state of *Hessen*, we investigate the persistence of the causal impact of age at school entry on secondary school track choice. Based on exogenous variation in the age at school entry created by month of birth, we obtain regression discontinuity estimates for different cohorts and grades up to the end of secondary education. We show that the age at school entry effect on grammar school attendance disappears exactly at the grade level where educational institutions facilitate track modification.

JEL classification: I21, I28, J24

Keywords: education, policy, identification, instrumental variables, relative maturity

Acknowledgement: This research was partly supported by the Anglo-German Foundation within the project ‘The Economics and Politics of Employment, Migration and Social Justice’, which is part of the Foundation's research initiative "Creating Sustainable Growth in Europe". We are grateful to Christian Dustmann, Horst Entorf, Bernd Fitzenberger, Hendrik Jürges, Michael Lechner, Stephen Machin, Sandra McNally, Regina Riphahn, Claus Schnabel, Reinhold Schnabel, Kerstin Schneider, Uta Schönberg and seminar participants of the Population Economics and Economics of Education groups of the German Economic Association for helpful comments. We also thank Hans-Peter Hafner of the Research Data Center (*Forschungsdatenzentrum*) of the Statistical Office of the state of *Hessen* for help with the administrative data for *Hessen*. Björn Schumacher and Philip Yang provided excellent research assistance. All remaining errors are our own.

Table of Contents for Chapter 3

3.1 Introduction to Chapter 3

The importance of path-dependencies of human capital investments is especially obvious in sports: If young players *within a cohort* are tracked into training groups by relative performance, this is known to generate an advantage for relatively older and thus physically more developed players who are more likely to reach the highest level group. As a consequence, relatively older players consecutively obtain training of higher quality so that the younger players within the same cohort fall behind (Allen and Barnsley, 1993).

In a more general context, recent education research considers early human capital investments to be crucial in light of dynamic self-productivity and complementarities of acquired skills and abilities. The survey article by Cunha *et al.* (2006) thus demonstrates that early human capital investments in disadvantaged children yield higher returns than later investments.

In this paper, we empirically analyse a specific application where educational institutions create path dependency but allow for later corrections of the initially chosen track. First of all, we consider how students' age at school entry (which is related to their relative maturity) affects their outcomes in secondary school. Secondly, we observe how this effect interacts with an early and rigorous secondary school tracking regime which mainly facilitates track revision six years after initial track choice.

Although almost all school systems in industrialised countries have some kind of ability grouping or tracking, Germany is one of the few countries that physically segregates students after grade 4 (at around the age of 10) into an academic and a non-academic track (see Brunello and Chechhi, 2006, for an overview of tracking systems). We show that age at school entry has a causal impact on track choice in grade 5. However, this effect disappears six years later (in grade 11) due to the possibility of track revision. Track revision mostly occurs through track upgrading, which is to a large extent facilitated by particular grammar schools that have historically emerged from Germany's vocational education system. Thus, unlike previous papers on the age at school entry effect, we show how the *institutional design* of the tracking regime generates and – once track choice can be revised – eliminates the age at school entry effect on track choice.

In the seminal paper by Angrist and Krueger (1992), the authors exploit a particular feature of the American school system by which quarter of birth generates exogenous variation in the duration of mandatory schooling: Because in the U.S. compulsory schooling ends at age 16, students born earlier in the year enter school at a later age and thus have a shorter duration of compulsory schooling. Based on census data und using quarter of birth as an instrumental variable for age at school entry, Angrist and Krueger (1992) show that age at school entry affects educational attainment. However, once the authors condition on school attendance beyond compulsory education, the age at school entry effect disappears. This finding is consistent with the interpretation that age at school entry only influences potential school dropouts in the U.S. institutional context. The importance of relative age effects, *i.e.* disadvantages caused by the fact that some students within the same class are younger compared to their peers seems to disappear over time.

Following this study, several methodologically similar papers have been published recently referring to a wide set of countries. Bedard and Dhuey (2006) use data for 18 industrialised countries from the international TIMSS study, as well as additional data for the US and Canada to estimate the effect of age at school entry at grades four and eight, where month of birth acts as an instrumental variable. Bedard and Dhuey (2006) establish that age at school entry has a positive impact on test scores at grades 4 and 8 in (virtually) all industrialised countries investigated. Although this effect is slightly smaller in grade 8 than in grade 4, the authors still find that a 11 months age difference increases the probability of college enrolment in the United States by 11 percentage points. These results contrast with those found by Angrist and Krueger (1992) for the U.S. Note that in many countries, month of birth does not influence compulsory schooling, so that the results of Bedard and Dhuey (2006) suggest that there is a persistent (relative) age effect on educational outcomes. Similarly, Allen and Barnsley (1993) show how quarter of birth (determining age at school entry) affects schooling outcomes in Britain and Canada. Although the effects seem to decline with progressing school grades, the effects persist until grade 9, when students are about 16 years of age. In particular, the authors show that quarter of birth has an impact on track attendance in British schools. For Sweden, Fredriksson and Öckert (2006) show that starting school later has a positive effect on educational outcomes and on earnings later in live. However, the size of the effect diminishes over the life cycle and earnings effects of a later school starting age are even negative when the opportunity cost of lost earnings of starting school (and hence working life) late are taken into account. By providing separate estimates of the age at school entry effect for several birth cohorts, the authors show that the age at school entry effect on years of schooling was larger for earlier cohorts, who still experienced a tracking system at school (similar to Germany today), which was replaced in the late 1960s with a comprehensive school system.

Although Fredriksson and Öckert (2006) do not dwell on the relationship between tracking and the age at school entry effect, their study and the one related to sports by Allen and Barnsley (1993) are – to the best of our knowledge – the only studies which provide evidence on the importance of tracking institutions on the age at school entry effect.[55] By focussing explicitly on track choice and by following several cohorts of students across different stages of secondary schooling, our study provides more explicit evidence on the impact of the tracking design on educational outcomes.

Early school tracking has been criticized to generate or perpetuate inequality: Using variation in tracking across countries, Brunello and Checchi (2006) find that early tracking reinforces initial differences between students, like the family background effect. Dustmann (2004) illustrates how the German regime of early tracking generates high intergenerational correlation of track choice. Hanushek and Wößmann (2006) show how countries with tracking increase the variance of student test scores (without raising the mean) relative to countries without tracking. Similarly, Bauer and Riphahn (2006) exploit variation in tracking age between Swiss cantons (administrative regions) and show that late tracking decreases inequality between students of high versus medium social background. Meghir and Palme (2005) evaluate a policy reform in Sweden which involved the abolition of school tracking and find that earnings of persons with high-skilled fathers fell but those of persons with low-skilled fathers rose because of the reform (which, apart from de-tracking also increased the duration of compulsory schooling and changed the curriculum). In contrast to these studies, Goux and Maurin (2006), find no evidence that de-tracking (easier access to grammar school) in Northern Ireland had a differential impact on students with different social backgrounds. Pekkarinen (2005) even finds that later tracking (at age 15 rather than 10) hurts boys with non-academic social background and relates this result to the timing of puberty.

None of the previous studies on tracking has explicitly focussed on the possibility to correct the tracking decision during the later years of secondary schooling. It is, however, an interesting feature of the German tracking system that it allows revising the tracking decision,

[55] In Puhani and Weber (2007) as in Chapter 2 of this thesis we show that age at school entry influences test scores at the end of elementary school as well as track choice in the middle of secondary school. However, at the time of writing, the administrative data on vocational schools were not available yet so that track change after grade 10, which turns out to be a crucial feature of the tracking system, could not be analysed. Neither were we able to follow cohorts over time, because only one cross section of administrative data on general schools was available.

especially after grade 10. This revision can go either way because grammar school students have the option to downgrade by entering an apprenticeship and students from lower track schools may upgrade to a grammar school.

In the present paper, we use administrative data in order to observe six entire school entry cohorts in a major German state for five consecutive years. In order to answer the question whether the age at school entry effect vanishes by the end of the high school period (grade 13 in Germany), we need to observe students in both general and vocational schools, since the grammar school degree (*Abitur*) can also be obtained in so-called 'vocational grammar schools' (*berufliches Gymnasium, Fachoberschule*). Because of this data requirement, we focus on analysing effects for students in the state of *Hessen*, which is the only state that made person-level data on pupils in general and in vocational schools with all the required information available to us.[56]

The paper is structured as follows: Section 3.2 describes the administrative data sets for the state of *Hessen* which are used in the following. Starting from stylized facts calculated from this data base, we also present a short overview of key institutional features of the school system in *Hessen* as compared to Germany as a whole. Specific emphasis is given to observed transitions between school tracks, which are crucial to the present analysis. Section 3.3 outlines the empirical strategy to estimate the causal impact of age at school entry on track attendance. Estimation results are presented and discussed in Section 3.4. We show that a later age at school entry has a persistent effect on track attendance up to grade 10. However, the possibility to revise track choice at grade 11 eliminates the age at school entry effect. We show that institutional design rather than time spent in school is responsible for the elimination of the age at school entry effect. No systematic differences are found between boys and girls. Section 3.5 concludes.

---

[56] *Hessen*, which includes the city of Frankfurt, is one of the larger German states. In terms of population size it is the fifth largest among the 16 German states and in terms of area it is the seventh largest.

3.2 Institutional Facts and the Administrative Data Source

3.2.1 School Tracking

Generally, tracking in Germany means that students are streamed into three types of secondary school after fourth grade (at about the age of ten), *i.e.* at a relatively early point of their educational career. Unlike in the U.S., tracking in Germany implies physical segregation of students into different schools. The most able students are supposed to attend the *Gymnasium* (comparable to the traditional British ‘grammar school’), which is the highest secondary track. The *Gymnasium* lasts for nine years and prepares for tertiary studies at academic institutions like three or four year colleges and universities (the equivalent of former British ‘polytechnics’ and ‘universities’).[57] As an alternative to *Gymnasium*, intermediate and lower level secondary school tracks lasting six or five years are provided by so-called *Realschule* and *Hauptschule*, respectively. Education at these schools is supposed to be less academic and more vocational compared to *Gymnasium* and typically prepares students for apprenticeships which imply continued part-time secondary education at vocational schools (*Berufsschule*). In the following, we will only distinguish between grammar school (*Gymnasium*) and no grammar school, because we consider the differences between the vocational tracks *Realschule* and *Hauptschule* to be small: Students in the lowest track (*Hauptschule*) may simply stay on another year to obtain the same degree as students in the intermediate track (*Realschule*). There are even recent tendencies to combine the lower two tracks.

Conceptually, a student’s ability and elementary school performance are supposed to determine the choice of the secondary school track. In practice however, parents mainly decide on their children’s educational pathway. In sum, the German tracking system seems to produce low intergenerational mobility (Dustmann, 2004).

Beside the three traditional school tracks, comprehensive schools exist, too. In order to access the importance of the different school tracks, displays shares of different school types attended in grade 8 during the school year 2005/2006. It is shown that the shares of the three major tracks range between one fifth (lower secondary, *Hauptschule*) and one third (grammar school, *Gymnasium*) while about 15 percent of all German students are in comprehensive schools (*Gesamtschule*). As can also be seen from Table 3.1, the distribution of different

[57] Recently, there is a tendency to shorten *Gymnasium* duration to eight years.

school tracks in the West German state of *Hessen* (on which we focus in this study) is representative for the pattern we observe for Germany as a whole. Compared to the average *West* German state, comprehensive schools have a long tradition in *Hessen* so that its proportion of comprehensive school graduates is relatively high (15, 15 and 9 percent in *Hessen*, Germany, and West Germany, respectively). More specifically, there are two different types of comprehensive schools in *Hessen*: While so-called 'integrated' comprehensive schools (*integrierte Gesamtschule*) are really comprehensive (non-tracked), 'cooperative' comprehensive schools (*kooperative Gesamtschule*) track students *within* the school. In the latter case, all tracks are offered within one school and students may easily switch tracks.

Not only because of the tradition of comprehensive schools, but also for another institutional detail, the secondary school system in *Hessen* is known to be rather flexible: Some schools in *Hessen* offer so-called 'support stages' (*Förderstufe*) providing comprehensive education during grade five and grade six thus delaying tracking for two years. Hence, children are given two more years to mature to reach an appropriate tracking decision. According to own calculations from administrative data, nearly 30 percent of all fifth graders attend these 'support stage' (later tracking) schools in *Hessen*.

Most importantly, besides the mentioned comprehensive institutions, there are two more sources of flexibility in the tracking system. First of all, students may generally change tracks after initial track selection. Track modification may, according to the School Law, occur in all grades and all types of secondary school but is complicated in practice, because of different curricula by school types and the fact that the sending school needs to agree. Secondly, however, further potential flexibility within the tracking system is provided by the fact that students may correct their initial track choice by deciding to continue education at a traditional (*Gymnasium*) or vocational grammar schools (*berufliches Gymnasium, Fachoberschule*) after having graduated from a lower or intermediate secondary school. Although the term 'vocational grammar school' may sound contradictory to the reader, this name can be explained by Germany's educational traditions. The German vocational education system, stemming in parts from the medieval guild and inning system, has already since the Bismarck period experienced an expansion by the creation of 'higher' technical (engineering) and business administration schools, along with Germany's rise to an industrial power. Other specializations like agricultural studies, social studies, health, or nutrition followed. Some of these schools received the status of colleges (*Fachhochschule*) and technical universities (*Technische Universität*) in the 1960s and 1970s. It only fitted into this logic that since the 1960s, the vocational stream developed its own grammar schools whose degrees (*Abitur*) are now in most cases formally equivalent to grammar school degrees from traditional schools. In Germany's state-run university system, where universities usually cannot select their students by their own criteria, 'formal' equivalence is crucial.

3.2.2 Administrative Student-Level Data for the State of *Hessen*

In Germany, the states are responsible for the school system and therefore also for collecting administrative data on schools. Student level data which is of special interest to empirical researchers has only recently been collected in most states.[58][59]

The present paper is based on available administrative school data for the state of *Hessen*. Two different data sources are joined for the purpose of this study: The first data set includes the population of students in *general* (primary and secondary) education. The second data set contains the population of students in *vocational* (secondary) education. The latter data set is important because it registers graduates from the lower and intermediate level schools. This includes students who continue secondary education by attending vocational grammar schools like *berufliches Gymnasium* or *Fachoberschule*, where they obtain the college/university entrance certificate (*Abitur*), as well as students in standard vocational schools (for example students in apprenticeships, who have to spent about 2 days per week in a vocational school called *Berufsschule*). Hence, both data sources are needed to observe the universe of students in secondary education in the state of *Hessen*. The present study is (to our knowledge) the first research paper drawing on the newly available vocational school information.[60] Currently, five waves of both sets of administrative data, referring to the school years 2002/2003 through 2006/2007, are available. With the two data sets joined, we observe the population, not a sample, of students in secondary education in *Hessen*.

The information contained in the administrative data for the state of *Hessen* is exceptionally valuable for analyzing the effect of age at school entry until the end of secondary schooling: The available data waves can be used to examine the development of

[58] Except of the state of *Thüringen*, which started to gather some individual level information in general primary and secondary schools in 1992, all the other states did not collect such data before 2002 (or even later in most cases).

[59] In 2006, we contacted the states' statistical offices but where denied access to the data in all states except of *Hessen*. Currently, more states make administrative data on the general education system available. However, at the time of writing, the data on students in vocational schools are only made available for research by the state of *Hessen*.

age at school entry effects over school entry cohorts and grades (as estimated on annually available individual information). In some cases, the grade information is not identified in the data on vocational schools. It is also completely missing on vocational schools in the 2002/2003 data. We therefore group students by school entry cohort and follow these cohorts over time. This is equivalent to following cohorts across grades if students do not repeat or skip grades. An attempt to group students exactly by grades will be made later on in Section 3.4. The different cohorts and grades under consideration are summarized in Table 3.2. For example, (not considering grade repetitions and grade skipping) we observe the cohort of students entering school in 1998 in fifth grade in 2002/03 and up to ninth grade in 2006. Similarly, we observe students who started first grade in 1993 in tenth grade in the data wave of 2002 and may 'track' this cohort up to grade 13 (when part of the students are still in general schools and others are observed in vocational schools). Since all students in secondary education in the state of *Hessen* are covered by the data, the number of observations is relatively high and amounts to around 60,000 individuals per cohort and school year. Note, however, that persons who leave the school system drop out of our data set. Given that the school law of the state of *Hessen* requires students to attend at least 9 years of general schooling plus – for those who do not attend grammar school – 2 or 3 years of vocational schooling (depending on the length of the apprenticeship chosen), persons dropping out of the data before grade 13 will be students not attending the *Gymnasium* (grammar school) track. A typical example of a student dropping out of the data set is a person completing the lowest secondary track (*Hauptschule*) after grade nine and then completing a two-year apprenticeship after grade 11. A person may also drop out after grade 10 if he or she is not doing an apprenticeship.

Hence, the fact that some students are missing in the data for grade 11 and especially grade 12 and 13 is a result of track choice. When estimating the effect of age at school entry on track choice, dropping out of the data set is therefore an outcome: some (not all) students who do not attend *Gymnasium* (grammar school) will not be observed in grades 11 to 13. We will therefore add these observations back into our cross sections for grades 11 to 13 by way of simulation (see also Section 3.3.3).

The variables collected for each student are the grade level and school type, grade level and school type in the previous school year, the school district, gender, nationality, month and year of birth, and month and year of school entry. There is no person identifier

---

[60] The history of the educational system in Germany is the reason why these two administrative data sets are

across years, so we do not have panel data. However, due to the variable indicating the previous school type, we can retrospectively observe track changes.

Since combining the administrative data on general and vocational schools allows following cohorts in all educational tracks, the information on the previous track provides data on track modification behavior which may be crucial for determining the long-run educational effects of school starting age. Table 3.3 summarizes 'entry rates' to and 'exit rates' from the grammar school track. Note that grammar schools comprise general grammar schools (*Gymnasium*) as well as vocational grammar schools (*berufliches Gymnasium*, *Fachoberschule*). Entry rates are defined as the number of students entering grammar school (from a lower track level) in a given grade divided by the total number of students who were in grammar school in the previous grade.[61] Exit rates are defined as the number of students leaving grammar school in a given grade divided by the total number of students in grammar school in the previous grade.

For students in comprehensive schools, it may be hard to judge whether they can be categorized in the *Gymnasium* level category. However, students in *cooperative comprehensive schools* are observed in their respective track within school (see Section 3.2.1), which we assign accordingly. For students in *integrated comprehensive schools* (without streaming) we assume that they are not at the highest secondary level. In fact, information provided by the Federal Statistical Office indicates that only seven percent of students in integrated comprehensive schools attain a higher level degree, so that measuring them as lower track when in or below grade 10 seems justified.

Considering the 1998 school entry cohort between school years 2002/03 and 2003/04 in Table 3.3, for example, two percent of students previously in *Gymnasium* track decide to switch to a lower track while the entry rate amounts to nine percent.[62] As can also be seen in Table 3.3, switching rates are especially high between grades 6 and 7 (the entry rate is between 16 and 22 percent for the observed cohorts), which is due to the institution of 'support stages' allowing to defer tracking to age 12 instead of 10 in *Hessen*. Similarly, relatively high entry rates to the highest track level of 44 to 45 percent are observed for eleventh graders because graduates from the intermediate or lower level school tracks may decide to continue education at any type of grammar school (like *berufliches Gymnasium*) in

---

separate entities.

[61] This definition of the entry rate makes the difference between the entry and the exit rate equal to the rate of growth of students in the grammar school track. Note that, as defined, the entry rate may in theory exceed 1.

order to seek an *Abitur*-level degree allowing them to study at a higher education institution (three or four-year colleges or universities).[63] One reason why this figure is so high is that students in integrated comprehensive schools not leaving general schooling after grade 10 in order to do an apprenticeship may remain in school at the grammar school level and will then be counted as track upgraders in the data. However, only about one half of the high transition rate is driven by students in comprehensive schools. This shows that between grades 10 and 11, when students are about 16 years of age, the German tracking system is characterized by a very high degree of mobility, both due to comprehensive and grammar schools. The fact that grammar school entry rates also seem relatively high (at 17 percent) between 'grades' 11 and 12 is an effect of the grouping of students according to school entry year rather than actual grades attended. If we attempt to group students by grade attended (which is difficult due to partially missing information), we obtain grammar school entry rates of around 73 percent between grades 10 and 11 and only 6 percent between grades 11 and 12. Hence track upgrading seems related to the institutional flexibility of the school system after grade 10.

The central question we seek to answer is how the age at school entry causally affects track attendance as students progress through the tracked secondary school system in the German state of *Hessen*. In this context, it is particularly interesting whether the mobility between tracks is in some way related to the age at school entry. Before we discuss these questions empirically, the following section presents a regression discontinuity design approach to identify the causal effect of the age at school entry on track attendance.

3.3 Identification of Age at School Entry Effects on Track Attendance

3.3.1 Implications of The Hamburg Accord and Discretion in Track Choice

Similar to most other OECD countries, the school entry age is effectively assigned by law in Germany. Although each German state has its own School Law (*Schulgesetz*), there is a high degree of co-ordination between states. A result of this co-ordination has been the Hamburg Accord (*Hamburger Abkommen*), which is in force since 1964. According to the school entry age regulation of the Hamburg Accord children in all German states should start school in August of a given year if they turn six years of age before the end of June in that year.

[62] Entry rates may also include persons moving to *Hessen* from another German state. Given the information in our data it is not possible to distinguish between these movers and track changers.
[63] Graduates of the lowest track level (*Hauptschule*) have to take one year at the intermediate level (*Realschule*) first before attending a grammar school.

Children turning six in the second half of the calendar year, *i.e.* between July and December, are supposed to wait until the following year before entering school. The Hamburg Accord thus generates a variation in the assigned age at school entry between 6.08 and 7.08 years of age. For children born directly around the cut-off date (30th June/1st July) the assignment rule thus generates a difference in the school entry age of one year.

Although date of birth may generate random variation in the assigned age of school entry, the actual school entry age can deviate from the assigned age due to parents' discretion. In particular, article 58 of the School Law of the state of *Hessen explicitly* allows for a deviation from the assigned age: 'Children who turn 6 years of age after June 30th may enter school [in the same calendar year] by parents' application. The decision is made by the school principal with consideration of the school doctors' advice'.[64] However, it is not only possible to enter school at an earlier than the assigned age, but also at a later one: '(…) children, who do not have the required physical, intellectual or mental status of development for attending school may be held back from attending (…) school for one year by application of the parents or the school principal (…)' In addition, the law allows children to enter school later than assigned by the Hamburg Accord if their knowledge of German is insufficient. This decision can be made by the school principal after consulting the parents.

Figure 3.1 displays assigned and theoretical age of school entry, as well as *Gymnasium* (grammar school) attendance by month of birth for all school entry cohorts as observed during the school year 2005/2006. It is shown that a share of students deviates from their assigned age at school entry and that this deviation is larger the closer students are born to the cut-off date. Children born in June (who are supposed to enter school at the age of 6) tend to enter school later than assigned, whereas the opposite is true for children born in July (who are supposed to enter school at the age of 7). Nevertheless, the actual age at school entry jumps upward between months of birth June and July, albeit not to the same degree as assigned by the Hamburg Accord. The fact that *Gymnasium* (grammar school) attendance also jumps between June and July is suggestive that the age at school entry drives track choice, but a formal test will be carried out in Section 3.4.

Given the flexibility of the entry age regulation, it is not surprising that the distribution of actual school entry ages is wider than the one of assigned entry ages.[65] As exhibited in

[64] Here and in the following, we provide our own translation from the original German text.

[65] This flexibility exists in similar forms in virtually all school laws of German states. Only Berlin is somewhat less flexible in that it requires children who turn six years of age in a calendar year to enter school in that year (deferral is not possible in this case). Bavaria also plans to shift the cut-off date. These deviations from the

Figure 3.2, entering school a year earlier or later than assigned is not uncommon in *Hessen* (nor is it in other German states). As only the month and year, but not the day of birth is provided in the data, the assigned school entry age as measured varies between 6.17 and 7.08. In the cohorts we consider for the school year 2005/06, 21 percent entered school early, *i.e.* below the age of 6.17 and about 16 percent entered late, *i.e.* after the age of 7.08.

The explicit exceptions to the school entry rule suggest that the actual age at school entry is likely to be endogenous. This implies that even if month of birth and hence the *assigned* age at school entry were randomly assigned across children, the *actual* age at school entry must be expected to be correlated with the ability of a child. In particular, the regulations allow that comparatively less able or disadvantaged children enter school later. Similarly, ambitious parents can have their children enter school at a younger age if they convince the school principal. This is likely to cause more able children to enter school at a younger age than assigned by the Hamburg Accord.

The possibility of non-compliance with the Hamburg Accord makes it impossible to estimate the causal effect of the age at school entry on school track attendance by simple correlations or ordinary least squares regressions. In other words, if we define β_1 to be the causal effect of school entry age a on track attendance y, an ordinary least squares regression in the form of equation (8)

$$y_i = \beta_0 + \beta_1 a_i + \beta_2 x_i + \varepsilon_i \tag{8}$$

will lead to a biased estimate of β_1, because we have to expect that $E(\varepsilon | a) \neq 0$. More precisely, we expect the correlation between the age at school entry a and the unobserved skill component ε to be negative, generating downward bias in the estimate of β_1 due to reverse causation: less able children – who are also less likely to choose the grammar school track – tend to enter school at a later age.

Only if x were to contain all variables driving both the selection into early or late school entry and track attendance, an ordinary regression (or simple discrete choice model) of track attendance on age at school entry would be meaningful. In most applications, however, we cannot be certain whether we observe all these variables. The German administrative data for the state of *Hessen* does not provide many social-background variables nor does it contain

end-of-June cut-off are possible as the regulation in the Hamburg Accord has been made less binding after 1997.

any test score.[66] Therefore we cannot rely on OLS regression to estimate the causal effect of age at school entry on track attendance.

3.3.2 Exploiting the Exogenous Variation in the Assigned Age at School Entry

Because of the endogeneity of the actual age of school entry, we consider two instrumental variable strategies to estimate the effect of age at school entry on track attendance. First we may use the variation in the assigned age at school entry to construct a two-stage least squares estimator, where the assigned entry age acts as an instrument. More specifically, the Hamburg Accord generates an assigned school entry age z_i for each student *i*, based on the month of birth b_i of the student:

$$z_i = Z(b_i) = \begin{cases} \frac{(72+8)-b_i}{12} & if \;\; 1 \leq b_i \leq 6 \\ \frac{(84+8)-b_i}{12} & if \;\; 6 < b_i \leq 12 \end{cases}, \qquad (9)$$

where, for example, $(72+8)-b_i$ indicates the age at school entry (measured in months, as the day of birth is not observed) for children born between January and June.

The first stage of the two-stage least squares estimator regresses the actual age at school entry a_i on the exogenous instrument z_i and possibly other exogenous characteristics x_i. By definition of the ordinary least squares estimator, this first-stage regression splits the variation in the actual age at school entry into an exogenous component $\hat{a}_i$, driven by the Hamburg Accord assignment rule $Z(\cdot)$ (and other characteristics x_i) and into an endogenous component $\hat{\eta}_i$, which is uncorrelated with $\hat{a}_i$:

$$\begin{aligned} a_i &= \hat{\gamma}_0 + \hat{\gamma}_1 z_i + \hat{\gamma}_2 x_i + \hat{\eta}_i \\ &= \hat{a}_i + \hat{\eta}_i \end{aligned}. \qquad (10)$$

The estimated first-stage coefficient $\hat{\gamma}_1$ indicates the degree of compliance with the entry age rule assigned by the Hamburg Accord. The variation in the exogenous component of the observed school entry age is accordingly driven by the compliers with the assignment rule. As shown in Imbens and Angrist (1994), the instrumental variables estimator identifies the effect

[66] The administrative data for other states of Germany is identical in this respect.

of interest for the compliers with the assignment rule.[67] Equivalently, the second-stage regression of track attendance y on the exogenous component of the age at school entry $\hat{a}$ identifies the causal effect of age at school entry on track choice for the group of students who would change their age at school entry if their month of birth (and hence their assigned age at school entry defined by the Hamburg Accord) were hypothetically varied.

The question arises how representative this 'local average treatment effect' (Imbens and Angrist, 1994; *i.e.* the effect for compliers with the Hamburg Accord) is for the population as a whole. Empirically, this question cannot be answered without further assumptions, because the status of being a complier is not observed and the average treatment effect in the population is not identified without further assumptions.

The control function approach discussed in Garen (1984) and Card (2001) poses a random coefficients model and makes the assumption that the deviation of the age at school entry effect from the average treatment effect for an individual is a linear function of the residual of the first-stage equation, *i.e.* a linear function of the size of non-compliance. We consider this assumption as too strong for our application, since it involves the symmetry restriction that those who enter school too early would be harmed by late entry in the same way as those who enter school too late benefit from late entry. We have, however, produced control function estimates (not shown here) and found that they were only slightly larger (in some cases even identical) than the local average treatment effects presented in this paper. Hence, we consider the local average treatment effect to be an informative parameter.

Alternatively, we may compare estimates for cohorts with a higher compliance to cohorts with a lower compliance to obtain an idea for the direction of the bias. However, our results below will give no clear indication in this respect. From a theoretical point of view, one would reason that those who comply with the school entry rule are least affected by the age at school entry, so that the local average treatment effect is biased toward zero compared

[67] Imbens and Angrist (1994) consider the case where both the instrument and the impact variable are binary. In this case, $\hat{\gamma}_1$ would be a consistent estimate of the share of compliers in the population despite the fact that any single observation cannot be identified as a complier. In our application, the situation is slightly more complicated because each person can vary the age at school entry discretely by one or more years in either direction. Thus, the estimate $\hat{\gamma}_1$ is also influenced by students who would generally enter school too early, but would enter at the age of 5 (instead of 6) if born in June, but at the age of 6 (instead of 7) if born in July. As in total only 6 percent of students enter school very early (at the age of 5) or very late (at the age of 8), we expect $\hat{\gamma}_1$ to be roughly equal to the share of compliers with the Hamburg Accord in the population of June- or July-born children.

to the average treatment effect.[68] Finally, it is worth noting that, although one may argue about the interpretation of local average treatment effects, we are not so much interested in the absolut size of the effect but rather in the direction of change over time (as students move to higher grade levels).

The second-stage estimator can be obtained by ordinary least squares on the following estimating equation:[69]

$$y_i = \beta_0 + \beta_1 \hat{a}_i + \beta_2 x_i + \underbrace{\left(\varepsilon_i + \beta_1 \hat{\eta}_i\right)}_{\tilde{\varepsilon}_i}. \qquad (11)$$

If month of birth and therefore the instrument z_i are completely random, no control variables x_i are required. Including control variables may even make the two-stage least squares estimator inconsistent if they are not exogenous. On the other hand, exogenous controls improve the precision of the estimator. We therefore produce estimates with and without control variables as a robustness check. The control variables are gender, regional (county), and citizenship indicators.

Because we cannot test whether month of birth and hence our instrument is truly randomly assigned, a second possible identification strategy is a so-called ‘fuzzy regression discontinuity design’ (Hahn, Todd and Van der Klaauw, 2001). Technically, this procedure amounts to applying the same two-stage least squares estimator as just described, but only to the population of students born close to the cut-off date (*i.e.* in June or July). Table 3.4 and Table 3A.1 in the Appendix to Chapter 3 present the correlations between the instrument and the control variables gender, region (county) and country of citizenship for two selected cohorts in all five observed school years for the discontinuity (June- or July-born) and full population, respectively. It is shown that the correlations are either zero or close to zero and at most 0.03 in absolute value. This is a tentative indication that month of birth (which is driving the instrument) is random, even if we observe only very few socio-economic characteristics in the administrative data. Nevertheless, in order to address any remaining doubts about the exogeneity of the instrument (defined on the basis of month of birth), we will in the following focus on the population of students born in June or July. This identification strategy is more convincing in terms of isolating other factors that might correlate with season of birth and

[68] This need not be true, however, if the effect of the age at school entry has opposite signs within the population of non-compliers. In this case, the direction of bias is undetermined.

[69] We use the two-stage least squares procedure in ‘Stata’ to obtain correct standard errors.

track choice. Results for the full population of students (born in any month of the year) will be provided in the Appendix.

3.3.3 First-Stage Regressions

Coefficients of the first-stage regressions for the population of students born in June or July are displayed in Table 3.5. Here and in the following section, we only show the specifications without control variables, because estimates with and without control variables are almost identical. As mentioned in Section 3.2.2, a subpopulation of those students who do not choose the *Gymnasium* track drops out of the data set at grades 11, 12 or 13 depending on the type of school or apprenticeship chosen. Therefore, we simulate missing observations in grades 11, 12 and 13, such that the number of observations is identical from grades 10 to 13. Missing observations are always allocated to the lower track (*i.e.* not in grammar school), because they would be in the data if they seeked a higher secondary degree.[70] The allocation of the simulated observations to birth months June or July and school entry ages 6 or 7 is based on the number of missing observations in these cells in grades 11 through 13 relative to the number of observations in grade 10.[71] In order to document results based only on observed students (*i.e.* without simulated observations), the lower panel of Table 3.5 (as well as of the following tables) displays the corresponding estimates in the non-shaded area of the table. Note, however, that we expect these estimates to be biased, because dropping out of the data set is not random, but a result of having chosen a lower track.

Estimates of the first-stages of two-stage least squares regression are provided by cohort and school year together with the F-statistics to test for potential weak instrument problems, indicated as a rule of thumb if the F-statistic is below 10 (Staiger and Stock, 1997; Stock, Wright and Yogo, 2002). If the cohorts do not change much over time, e.g. through

[70] The fall in the number of students might in theory also be due to net out-migration from the state of *Hessen*. However, we find this explanation for the fall in the number of observations from grade 11 onwards implausible, as the published net-migration rate for the age group 6-18 years in *Hessen* is -1 percent and +12 percent for ages 18-25. We lose about 10 percent of the observations between grades 11 and 12. As students usually still live with their parents at these ages, it is unlikely that this loss of observations has something to do with educational migration. It is also unclear what students in *Hessen* would gain by switching state borders in grades 12 and 13 (when again we lose some observations, most likely due to apprenticeship completion). The high net-immigration rate (over 12 percent) for ages 18-25 is likely to be driven by college/university students and young workers. Migration by parents out of the state of *Hessen* cannot explain the facts, either, as it would occur more evenly across cohorts in a given calendar year and because net-migration of age groups 25-40 and 40-50 is close to zero or -2 percent, respectively. As it is evident that the loss of observations is related to the grade (or time in school) of the student, school dropouts and apprenticeship completion are the only possible explanation for the loss of observations.

[71] For each birth month, the proportions of students entering earlier or later than the theoretical age at school entry are held constant (relative to grade 10). For the simulated observations we assume that pupils entering earlier (later) compared to their theoretical entry age always enter one year earlier (later).

migration across state borders (see footnote 70), the first-stage coefficients should not vary much by school year within cohorts. However, first-stage coefficients might vary across cohorts if compliance behaviour changes from year to year. Considering the first column of Table 3.5 (referring to the school year 2002/03), there is indeed some variation in the degree of compliance across cohorts. The 1995 school entry cohort shows the lowest compliance with a first-stage coefficient of 0.31 and the 1997 school entry cohort the highest with a coefficient of 0.41. The ‘opt-out’ clauses in the German school laws explain why the assigned entry age is observed to a much lower extent in Germany than in Scandinavian countries like Sweden and Norway (Fredriksson and Öckert, 2006, and Strøm, 2004). However, none of the first-stage F-statistics point to a weak instrument problem. In the full population (Table 3A.2 in the Appendix to Chapter 3) the degree of compliance is mostly somewhat higher than in the discontinuity samples. Differences arise due to the behaviour of students born in months other than June or July.

Considering the first-stage estimates within cohorts across school years, variations in the point estimates are mostly minor (as expected). In addition, the number of observations per cohort usually does not vary by more than a percentage point.

3.4 School Entry Age Effects on Track Attendance in Secondary School

3.4.1 Ordinary Least Squares Regressions

As argued in Section 3.3.1, we expect the correlation between age at school entry and track choice to be driven by an overlap of both the causal effect of school entry age on track choice and a selection effect. Table 3.6 presents bivariate OLS estimates of *Gymnasium* (grammar school) attendance on the age at school entry. Separate estimates are provided by school entry cohort and school year.

As the table shows, from grade 6 to grade 10 (the grey-shaded areas) all estimates are negative and statistically significant without exception.[72] Hence, the correlation between age

[72] For the 1998 school entry cohort in 2002/03 (supposed to attend grade 5), the estimate is comparatively small in absolute value. The reason is that some students are still in primary school. These are both students who repeated a grade and students who entered school through a special type of pre-school (‘grade 0’, *Vorklassen*). This pre-school only exists in areas with disadvantaged backgrounds and effectively leads to a 5-year (instead of 4) elementary school period. Because we group students by school entry year, children who entered pre-school (‘grade 0’) in 1998 are still in elementary school (*i.e.* in fourth grade) when we observe them in 2004/2005, which implies that they cannot be observed in the grammar school track. This causes upward bias in the OLS estimate, as some young starters (the school entry age is registered irrespective of whether ‘grade 0’ or grade 1 are entered) are not in *Gymnasium* when surveyed in the school year 2002/2003.

at school entry and grammar school (*Gymnasium*) attendance is unambiguously negative in the middle of secondary school. A representative estimate indicates that those students who enter school at the age of 7 have a probability of attending grammar school that is 11 percentage points lower than the one of students entering school at the age of 6. Including gender, regional and country of origin controls into the regression only leads to small changes in the estimates (not shown here), with a tendency for the point estimate to become smaller in absolute value. This decrease in the absolute value of the OLS estimate is indicative of a correlation between the actual age at school entry and socio-economic characteristics that leads to a downward bias in the OLS coefficients, as argued in Section 3.3.1.

The OLS estimates are markedly different, at between zero and -4 percentage points, at grade 11.[73] Note that this is the transition where upward mobility into the *Gymnasium* track is particularly high according to Table 3.3. It seems that this mobility is affecting the correlation between the age at school entry and track attendance in the sense that late entrants, who are likely to be a selected group of students with less innate ability, make it into *Gymnasium* track at grade 11. Note that the age at school entry regulations allow students to defer school entry if they do not have the required level of development (Section 3.3.1). Assuming that these students have less innate ability, their deferral will generate a negative correlation between the age at school entry and *Gymnasium* attendance that is *not* causal. The fact that this correlation disappears at grade 11 suggests that track upgrading after grade 10 provides grammar school education to less talented students. This is an argument often made by conservative political circles who want to preserve rigorous tracking against institutional flexibility. The counterargument by the political left in favour of institutional flexibility is that track upgrading helps students with innate ability but disadvantaged backgrounds. As we will demonstrate below when analysing the *causal* effect of age at school entry, the eased possibility of track revision not only opens the gates of grammar schools to less talented students (as suggested by the OLS results), but also corrects the age at school entry effects. Hence, it promotes *some* students who are *able* (there should be no systematic innate

Excluding all those in elementary school changes the OLS estimate to -0.13, *i.e.* a similar value as observed for the other grades up to grade 10. Since the data do not record whether a student entered school through pre-school (about 4 percent do), we are not able to handle this issue directly. Because school entry into pre-school distorts the allocation of students to grades, we will try to group students by actual grades attended in Section 4.3.

[73] The coefficients turn significantly negative again 12 and especially 13 years after school entry. This may be related to the fact that some grammar schools (*Fachoberschule*) finish after grade 12 so that we make a mistake by assigning *all* simulated observations to a lower track. Note, however, that our main results below are not affected by this problem, as the change in the regression discontinuity estimates between grades 12 and 13 are not relevant for the story of the paper.

differences between June and July-born children), but have been disadvantaged by their relatively young age at the time of initial track selection at the age of 10. This latter effect can only be detected by causal analysis like instrumental variable estimation, which we now turn to.

3.4.2 Causal Effects: Regression Discontinuity Estimates

The regression discontinuity design estimates (two-stage least squares based on the population of students born in June or July) are provided in Table 3.7. It is remarkable that, in contrast to the OLS estimates, all estimates for up to grade 10 (grey shaded areas) are positive and different from zero in terms of statistical significance. The negative OLS estimates are therefore heavily affected by reverse causation and strongly downward biased.

Regression discontinuity design point estimates up to grade 10 range between 0.08 and 0.19, but the variation of the estimates is larger between than within cohorts (the range is between 0.11 and 0.16 for the 1998 school entry cohort and between 0.08 and 0.10 for the 1997 school entry cohort, for example). The median estimate in the grey-shaded region (grades 5 to 10) is 0.13 implying that the causal effect of entering school at age 7 instead of 6 is an increase in the probability to attend grammar school of 13 percentage points, which is large given that slightly over a third of all students attend grammar school (this effect comes into full force for a complier with the Hamburg Accord if his or her birthday is changed from June 30th to July 1st).[74] The standard deviations of the estimates lie between 2 and 3 percentage points. Additional control variables change the point estimates only slightly and in all cases by less than a standard deviation of any estimate. The corresponding estimates based on the full population of students (Table 3A.4 in the Appendix) are mostly a few percentage points higher than in the population of students born in June or July (Table 3.7).

Considering the two grade transitions when the tracking system exhibits the largest mobility, *i.e.* from grades 6 to 7 and from grades 10 to 11, a clear pattern emerges. First, the existence of the 'support stage' (*Förderstufe*) in some schools in the state of *Hessen* does not lead to a distinct change in the point estimate of the age at school entry effect between grades 6 and 7 (*cf.* the estimates for the 1997 and 1998 school entry cohorts in Table 3.7). Hence, the institutional mobility offered by these 'support stages' in the form of a deferred track choice at the age of 12 instead of 10 does not attenuate the age at school entry effect on track choice.

[74] Estimates at the mean obtained from probit instrumental variable models are almost numerically identical to the linear probability model estimates presented here.

By contrast, the possibility to correct the tracking decision after grade 10 has large consequences. None of the point estimates of the age at school entry effect are significantly different from zero for students having attained 12 years of schooling and only two out of four are significant when students have attained 11 years of schooling. Indeed, the drops in the point estimates between 10 and 11 years of schooling are very large and range between 5 and 13 percentage points, depending on the school entry cohort.[75] In the full population of students (Table 3A.4 in the Appendix), the results are very similar: point estimates drop between 6 and 13 percentage points between 10 and 11 years of schooling, and 12 years after school entry there are no significantly positive effects any more.

Although we do not have panel data, we observe the track attended in the previous school year for (almost) all students. As Table 3.3 has shown, changes of the school track are mostly upgrades to grammar school and occur predominantly after 10 years of schooling. In order to document the causal effect of age at school entry on track upgrade to grammar school directly, we present two-stage least squares estimates with track upgrade as the outcome variable in Table 3.8 (based on the population of students born in June or July).

Thus, Table 3.8 illustrates how the German tracking system is more likely to allocate students who enter school at a relatively older age to the grammar school track after elementary school and does not reverse this decision until six years later: The regression discontinuity design estimates referring to 5 years of schooling (the upper left dark-shaded figure) suggest that entering school at the age of 7 instead of 6 increases the probability to enter *Gymnasium* in grade 5 (when tracking starts) by 13 percentage points. As must be expected, these estimates correspond to the ones for track attendance in Table 3.7. In the following grades (from 6 to 10), the age at school entry has hardly any effect on track upgrading: the point estimates are close to zero (2 percentage points at the maximum) and often insignificant. This is not surprising as differences in curricula and other requirements make it difficult to change tracks during the middle of secondary school (see Section 3.2.1). This is different when students enter their eleventh year of schooling: At this stage, graduates from the schools below grammar school (*Gymnasium*) track have to decide whether to enter apprenticeship training (with ordinary vocational schooling) or to move to a traditional or

[75] If we do not simulate the observations that we lose in grades 11 to 13 due to school dropouts and completed apprenticeships, we still observe a large drop in the point estimates, although not as large as when the lost observations (which are a result of track choice) are taken account of (see the lower part of Table 3.7).

vocational grammar school. This is the time when track upgrading is facilitated by the German tracking system.

The estimates in Table 3.8 show that, 11 years after school entry, track upgrading is causally influenced by the age at school entry. Students who entered school at a relatively older age are less likely to upgrade. In other words, students who entered school at a relatively young age (6 instead of 7) are more likely to upgrade. The point estimates indicate that entering school at the age of 7 instead of 6 decreases the probability to upgrade to the grammar school track by between -4 and -8 percentage points. A year later, 12 years after school entry, the effect is still between 0 und -3 percentage points, adding up to an effect between -6 and -8 percentage points in each cohort.[76] Comparing effects of age at school entry on track attendance and track upgrade in Table 3.7 and Table 3.8 shows that track upgrading explains – depending on the cohort – more than half or almost all of the disappearance of the age at school entry effect on track attendance. For the school entry cohorts 1995 (cohort 4) and 1993 (cohort 6) a later age at school entry also has a significant effect on track downgrade (results not shown here), which – together with the results on track upgrade in Table 3.8 – explains the size of the declines in the estimates in Table 3.7.

Given these significant effects of age at school entry on track attendance until 10 or 11 years after school entry and their elimination afterwards, one might wonder whether these effects are different for boys and girls. It turns out that compliance with the Hamburg Accord’s rule on the age at school entry is very similar for boys and girls (first-stage coefficients, not shown here, hardly vary between the genders). In the Appendix (Table 3A.6 and Table 3A.7), we present two-stage least squares coefficients for June or July born men and women, respectively. No systematic differences arise between boys and girls. For both genders, there is a significant age at school entry effect until 10 years after school entry which becomes insignificant 12 years after school entry at the latest.

[76] The estimates without simulated missing observations (lower panel of Table 3.8) are lower in absolute value but qualitatively similar.

3.4.3 Is the Age at School Entry Effect Eliminated by Institutions or Time Itself?

We have shown in the previous section that relatively young students at school entry tend to move to lower tracks at about the age of 10 (when tracking starts), but tend to upgrade their track choice at the time when they attain their eleventh year of schooling. Although it is striking that mobility occurs exactly when track change is facilitated institutionally, it is not clear whether institutions or simply time in school itself are the cause for the systematic track upgrade by relatively young school entrants.

In order to check whether it is really the institutions that matter or rather time in school, we carry out two different checks. First, instead of grouping students by the school entry year (and hence the time spent in school), we attempt to group them by grades. We write 'we *attempt*' because the administrative data on vocational schools does not give direct information on the grade attended in all cases (it is missing entirely for the school year 2002/2003 as well as for certain types of schools in all school years), which is why we have stuck to grouping students by year of school entry for the main part of the paper. If it is institutions that matter rather than years spent in school, the reduction in the age at school entry effect should be larger between grades 10 and 11 when grouping students according to actual grade attended rather than year of school entry.

Second, we define an alternative outcome variable, by counting only traditional grammar schools as *Gymnasium* track and coding the vocational grammar schools (*berufliches Gymnasium* and *Fachoberschule*) as not attending *Gymnasium*. Comparison of the results based on this alternative outcome variable with those discussed in sub-section 3.4.2 reveals whether students are upgrading mainly through vocational or also through traditional grammar schools. Note that both types of grammar school award high school degrees that in most cases carry the same rights of entering tertiary educational institutions like colleges or universities (see Section 3.2.1 for more detail and Section 3.4.4 for a discussion of potential wage effects of attending different grammar schools). If track upgrading occurred mostly to vocational grammar schools rather than traditional ones, this would be another indication for the importance of institutions (here vocational grammar schools) in eliminating the age at school entry effect.

Table 3.9 reports the regression discontinuity design estimation results for the grouping according to grades. Note that we lose the first year (2002/2003), because grade information is not available for the administrative data on vocational schools for this year. In order to gauge whether institutions or years in the school system matter more, the estimates based on the grouping by grade have to be compared with those of the grouping by school entry cohort (Table 3.7). We can compare the change in the effect of the school entry age on grammar school attendance between 'grades' 10 and 11 for cohorts 3 to 5 (cohort 6 is only observed when students are already in grade 11 because we cannot use data on the school year 2002/03 when grouping students by grade). The comparison between Table 3.7 and Table 3.9 clearly shows that the fall in the age effect is much larger when grouping students according to the actual grade attended rather than years since school entry. Indeed, whereas the median estimate for 10th graders is still between 0.12 and 0.13, none of the estimates for grade 11 remain significant, with a maximum point estimate of 0.05. The point estimates for grades 11 through 13 are similar and insignificantly different from zero without exception. Note that grouping students according to the year of school entry results in a somewhat more protracted drop in the age at school entry effect, which is zero for all cohorts only 12 years after school entry (Table 3.7). However, in the light of the results of Table 3.8, where we group students by grade, it seems that the timing of the elimination of the age at school entry effect can be explained by some students repeating a grade during secondary school and thus attending grade 11 in their 12th year of schooling.[77] In sum, the findings substantiate the view that the age at school entry effect on track attendance is eliminated in grade 11 and that the institutional flexibility provided by the education system after grade 10 rather than personal maturity is responsible for the elimination of the relative age effect on track attendance.

Further support for this view is given by the estimation results shown in Table 3.10, where we only count traditional *Gymnasium* as grammar schools. Vocational grammar schools like *berufliches Gymnasium* and *Fachoberschule* are coded as lower track in order to check whether the elimination of the age at school entry effect occurs through traditional or more recently created vocational grammar schools. It turns out that both types of grammar schools are about equally important. In Table 3.10, the age at school entry effect remains significant until 13 years after school entry, which suggests that without vocational grammar schools, the age at school entry effect would not completely vanish *ceteris paribus*. This demonstrates that vocational grammar schools (*berufliches Gymnasium* and *Fachoberschule*) are important institutions for correcting the track decisions made at the end of elementary school. However, depending on the cohort under consideration, a comparison of the estimates in Table 3.7 and Table 3.10 reveals that more or less one half of the decline in the point

[77] Two-stage least squares estimates of the effect of age at school entry on grade repetition (not shown here) are very close to zero and statistically insignificant in virtually all cases.

estimate of the age at school entry effect is explained by traditional grammar schools. Hence both traditional and vocational institutions contribute to the elimination of the age at school entry effect.

3.4.4 Do We Expect Any Age at School Entry Effects on Wages?

Given that the data available to us measure only track attendance until the end of secondary school, the question arises what impact track attendance has on the labour market. The second question in this context is in what way different types of grammar schools are comparable economically rather than formally. Because available labour force surveys in Germany do not provide information on month of birth, we cannot take the direct route and estimate effects of month of birth on wages. Instead, we use the 2004 German Socio-Economic Panel (GSOEP) to estimate the return to different types of grammar school. We start by regressing log hourly wages on a dummy variable indicating any type of grammar school degree. Results are shown in Table 3.11. The only control variables are age and age squared (only West German workers aged 26 to 40 are included who attended school from the 1970s onwards, when the current German schooling system was already in place). For men, the estimated return to completing grammar school (potentially involving attending college/university, which is not controlled for in the regressions) amounts to 21 percent, for women, it is 24 percent. Similar regressions with higher education as the outcome variable (not shown here) suggest that grammar school completion raises the probability to obtain a college/university degree by 51 percent for men and 48 percent for women.

What are the returns to different types of grammar school? The GSOEP allows distinguishing between the degree of *Fachabitur* (one particular type of vocational grammar school degree, which in some cases restricts the subjects or institutions of higher education, obtained at *Fachoberschule*) and *Abitur* (which can be obtained at traditional grammar schools, *Gymnasium,* or vocational grammar schools, *berufliches Gymnasium*). We estimate similar wage regressions as above to test whether the labour market returns are different between the two groups of grammar school degrees (also shown in Table 3.11). For men, the difference in the return between *Abitur* (obtained at traditional *Gymnasium* or *berufliches Gymnasium*) is a statistically insignificant 3 percent. For women, it is a statistically insignificant -1 percent. Hence, combining all types of grammar schools into a single category as in the main analysis of the paper seems justified. Although we cannot determine the age at school entry effect on wages, these wage regressions at least suggest that the systematic revision of track choice that leads to the elimination of the age at school entry effect on track attendance after grade 10 might also eliminate any potential effects of age at school entry on wages emerging through different higher secondary institutions.

3.5 Conclusions of Chapter 3

Based on a regression discontinuity design, we provide evidence on the effect of age at school entry (relative maturity) on track attendance in secondary school using administrative data containing the population of students in the German state of *Hessen*. This is – to the best of our knowledge – the first time this administrative data on individual students from *both* general *and* vocational schools in a German state is used for empirical research. We follow six different cohorts across time for five school years.

Germany is characterized by a tracking system with physically separate types of secondary schools, where only the highest track (grammar school, *Gymnasium*) allows access to tertiary college/university education. Especially after grade 10, the German tracking system offers the possibility to revise the tracking decision made after grade 4, because students who have not attended grammar school may opt for a traditional or a so-called vocational grammar school, the latter of which offers a degree equivalent to the one of traditional grammar schools. We show that the tracking system generates a statistically significant effect of age at school entry on the probability to attend the highest track level (grammar school, *Gymnasium*) until 10 years after school entry: a relatively young age at school entry significantly decreases the probability to attend grammar school by about 13 percentage points. However, the age at school entry effect disappears 11 or 12 years after school entry, depending on the school entry cohort we consider.

Additionally, we show that the institution of vocational grammar schools is crucial by accounting for about one half of the elimination of the age at school entry effect (depending on the cohort considered). The other half is accounted for by traditional grammar schools accepting – after grade 10 – students who had previously attended lower tracks. For some cohorts, track downgrade also plays a role in the elimination of the age at school entry effect, because grammar school students may enter the vocational apprenticeship system after grade 10 without further requirements and without completing a higher level secondary degree. In general, however, track upgrading after grade 10, which – as the regression discontinuity design estimates show – is causally affected by the age at school entry, is key to the elimination of the age at school entry effect.

By grouping students both by the year at school entry and by grades attended, we show that the age at school entry effect is eliminated after grade 10 rather than simply after a certain number of years in school. Hence, the design of institutions seems to matter for whether and when the age at school entry effect is eliminated. This does not prove that years of schooling do not play a role at all: indeed, the institution of the ‘support stage’ (*Förderstufe*) which allows deferral of the tracking decision for two years, has no effect on the impact of age at school entry on the tracking decision.

In sum, our results give support to policies which facilitate a correction of the school tracking decision early enough before the decision whether to attend college/university education is made. Our estimates do not prove the existence of tracking in Germany as inefficient *per se.* Yet the fact that a large age at school entry effect on grammar school attendance exists during the first six years of secondary school seems to be an indication of misallocations generated by this system. We have shown that relatively young students – who lag behind their peers only in relative maturity, as birth in June or July can be treated as random – are systematically more likely not to be selected into grammar school. This strong result raises the question whether other characteristics that have – at least partly – only temporary effects on capabilities, like language problems of immigrant children or some types of parental influences on children’s early school performance, equally generate misallocations in a tracking system.

References for Chapter 3

Allen, J. and R. Barnsley (1993): Streams and Tiers, The Interaction of Ability, Maturity, and Training Systems with Age-Dependent Recursive Selection, *Journal of Human Resources* 28: 649-659.

Angrist, J.D. and A.B. Krueger (1992): The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples, *Journal of the American Statistical Association* 87: 328-335.

Bauer, Ph. and R.T. Riphahn (2006): Timing of School Tracking as a Determinant of Intergenerational Transmission of Education, *Economics Letters* 91: 90-97.

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

Brunello G. and D. Checchi (2006): *Does School Tracking Affect Equality of Opportunity? New International Evidence,* IZA Discussion Paper No. 2348.

Card, D. (2001): Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica* 69: 1127-1160.

Cunha, F., J.J. Heckman, L. Lochner and D.V. Masterov (2006): Interpreting the Evidence on Life Cycle Skill Formation, in E. Hanushek and F. Welch F (eds.): *Handbook of the Economics of Education*, Vol.1, Elsevier, Amsterdam: North-Holland, 697-812.

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

Fredriksson, P. and B. Öckert (2006): *Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labour Market Performance*, IFAU Working Paper No. 12.

Garen, J. (1984): The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable, *Econometrica* 52: 1199-1218.

Goux, D. and E. Maurin (2006): *Close Neighbours Matter: Neighbourhood Effects on Early Performance at School,* IZA Discussion Paper No. 2095.

Hahn, J., P. Todd and W. Van der Klaauw (2001): Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design, *Econometrica* 69: 201-209.

Hanushek, E.A. and L Wößmann (2006): Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries, *Economic Journal* 116: 63-76.

Imbens, G.W. and J.D. Angrist (1994): Identification and Estimation of Local Average Treatment Effects, *Econometrica* 62: 467-475.

Maurin, E. and S. McNally (2007): *Educational Effects of Widening Access to the Acadmic Track: A Natural Experiment*, IZA Discussion Paper No. 2596.

Meghir, C. and M. Palme (2005): Educational Reform, Ability, and Family Background, *American Economic Review* 95: 414-424.

Pekkarinen, T. (2005): *Gender Differences in Educational Attainment: Evidence on the Role of the Tracking Age from a Finnish Quasi-Experiment*, IZA Discussion Paper No. 1897.

Puhani, P.A. and A.M. Weber (2007a): Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany, *Empirical Economics* 32:359-386.

Staiger, D. and J.H. Stock (1997): Instrumental Variables Regression with Weak Instruments, *Econometrica* 65: 557-586.

Stock, J.H., J.H. Wright and M. Yogo (2002): Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics* 4: 518-529.

Strøm, B. (2004): *Student Achievement and Birthday Effects*, Norwegian University of Science and Technology, mimeo.

Tables and Figures for Chapter 3

Table 3.1: Attendance of German school tracks in grade 8 in 2005/2006 (Percentages)

	Germany	West-Germany	Hessen
Lower Secondary (*Hauptschule*)	22	26	17
Intermediate Secondary (*Realschule*)	26	29	27
Higher Secondary (*Gymnasium*)	31	31	35
Comprehensive Schools[A] (*Gesamtschule*)	15	9	15
Special Schools[B] (*Sonderschule*)	1	1	1
Free Waldorf-Schools (private)[C]	5	5	5

Note: [A]Comprehensive Schools include further combined school types. [B]For children with special needs, mostly due to physical or mental disabilities. [C]Waldorf-Schools follow a special educational philosophy and may lead to different secondary degrees.
Source: Federal Statistical Office (2006), Fachserie 11.1.

Table 3.2: School grades in which school entry cohorts are observed

	(entry year)	2002/03	2003/04	2004/05	2005/06	2006/07
Cohort 1	(1998)	5	6	7	8	9
Cohort 2	(1997)	6	7	8	9	10
Cohort 3	(1996)	7	8	9	10	11
Cohort 4	(1995)	8	9	10	11	12
Cohort 5	(1994)	9	10	11	12	13
Cohort 6	(1993)	10	11	12	13	---

Note: Grades refer to the supposed grade levels of students (if grades are not repeated or skipped) who entered school in the indicated year and are observed between 2002/2003 and 2006/2007.

Table 3.3: Grammar school entry and exit rates

Entry rates	(entry year)	2002/03	2003/04	2004/05	2005/06	2006/07
Cohort 1	(1998)	---	0.09	**0.16**	0.02	0.01
Cohort 2	(1997)	0.13	**0.22**	0.02	0.01	0.01
Cohort 3	(1996)	**0.18**	0.04	0.01	0.01	**0.45**
Cohort 4	(1995)	0.03	0.01	0.01	**0.44**	0.17
Cohort 5	(1994)	0.01	0.02	**0.44**	0.17	0.08
Cohort 6	(1993)	0.01	**0.45**	0.17	0.08	---
Exit rates	(entry year)	2002/03	2003/04	2004/05	2005/06	2006/07
Cohort 1	(1998)	---	0.02	**0.02**	0.02	0.03
Cohort 2	(1997)	0.04	**0.02**	0.02	0.03	0.03
Cohort 3	(1996)	**0.03**	0.02	0.03	0.03	**0.02**
Cohort 4	(1995)	0.04	0.03	0.03	**0.03**	0.03
Cohort 5	(1994)	0.05	0.04	**0.03**	0.03	0.03
Cohort 6	(1993)	0.05	**0.03**	0.03	0.03	---

Note: The entry rates into grammar school (Gymnasium) are so high between grades 10 and 11, as they also capture students from comprehensive schools who stay on to obtain the university entry certificate (*Abitur*). Entry rates are defined as the ratio of students entering grammar school (from a lower track level) in a given grade related to the total number of students who had been in grammar school in the previous grade. Exit rates are defined as the number of students leaving the highest secondary track in a given grade divided by the total number of students in this track in the previous grade.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.4: Correlations between instrument and observables – Population of students born in June or July

	Variable	2002/03	2003/04	2004/05	2005/06	2006/07
Cohort 1						
(1998)	Male	-0.01	0.00	0.00	0.00	0.00
	Region 1 (Darmstadt)	0.01	0.01	0.00	0.00	0.01
	Region 2 (Frankfurt)	0.00	0.00	0.00	0.00	0.00
	Region 3 (Offenbach)	0.00	0.00	0.00	0.01	0.01
	Region 4 (Wiesbaden, Main-Taun., Rheing.)	0.00	0.01	0.01	0.00	0.00
	Region 5 (Bergstr., Odenw., Diebg., G.-Ger.)	0.00	0.00	0.01	0.01	0.01
	Region 6 (Hochtaunus, Wetterau)	0.01	0.01	0.01	0.01	0.01
	Region 7 (Main-Kinzig)	-0.01	-0.01	-0.01	0.00	-0.01
	Region 8 (Gießen, Lahn-Dill, Limburg-Weil.)	-0.01	-0.01	-0.01	-0.01	-0.01
	Region 9 (Marburg-Biedenkopf, Vogelsberg)	0.01	0.01	0.00	0.00	0.00
	Region 10 (Kassel)	-0.01	-0.01	0.00	-0.01	0.00
	Region 11 (Fulda, Hersfeld-Rotenburg)	-0.01	-0.01	-0.01	-0.01	-0.01
	Region 12 (Kassel-Land, Werra-Mei., Sch.-E)	0.00	0.00	0.00	0.00	0.00
	Nationaliy: German speaking country	0.01	0.01	0.01	0.01	0.01
	Nationaliy: Turkey	-0.01	-0.01	-0.01	-0.01	-0.02**
	Nationaliy: Italy, Greece	-0.01	0.00	-0.01	-0.01	-0.01
	Nationaliy: Former Yugoslavian states	-0.01	-0.01	0.00	0.01	0.01
	Nationaliy: Remaining 'Western' countries	0.00	0.00	-0.01	-0.01	-0.01
	Nationaliy: Eastern Europe, former Soviet Un.	-0.01	-0.01	0.00	0.00	0-00
	Nationaliy: Remaining 'Muslim' countries	0.01	0.00	0.00	0.00	0.01
	Nationaliy: Remaining Asia	0.00	0.01	0.01	0.01	0.01
	Nationaliy: Remaining countries	0.00	0.00	-0.01	-0.01	0.00
Cohort 6						
(1993)	Male	-0.02*	-0.03**	-0.01	-0.02	---
	Region 1 (Darmstadt)	0.01	0.00	-0.01	0.00	---
	Region 2 (Frankfurt)	0.01	0.01	0.00	0.00	---
	Region 3 (Offenbach)	0.00	-0.01	0.01	0.02*	---
	Region 4 (Wiesbaden, Main-Taun., Rheing.)	0.01	0.00	0.00	-0.01	---
	Region 5 (Bergstr., Odenw., Diebg., G.-Ger.)	0.00	0.00	-0.01	0.00	---
	Region 6 (Hochtaunus, Wetterau)	0.01	0.01	0.00	-0.01	---
	Region 7 (Main-Kinzig)	0.00	-0.01	-0.01	-0.01	---
	Region 8 (Gießen, Lahn-Dill, Limburg-Weil.)	-0.02	-0.01	0.00	0.00	---
	Region 9 (Marburg-Biedenkopf, Vogelsberg)	0.02*	0.01	0.00	0.01	---
	Region 10 (Kassel)	-0.03**	0.00	0.01	0.01	---
	Region 11 (Fulda, Hersfeld-Rotenburg)	0.00	-0.01	0.01	0.00	---
	Region 12 (Kassel-Land, Werra-Mei., Sch.-E)	0.00	0.01	0.00	-0.01	---
	Nationaliy: German speaking country	0.02	0.01	0.01	0.02	---
	Nationaliy: Turkey	-0.03**	-0.02**	-0.02	-0.02**	---
	Nationaliy: Italy, Greece	0.00	0.02*	0.01	0.01	---
	Nationaliy: Former Yugoslavian states	0.00	0.00	-0.01	-0.01	---
	Nationaliy: Remaining 'Western' countries	0.00	0.00	0.00	0.00	---
	Nationaliy: Eastern Europe, former Soviet Un.	0.00	-0.01	-0.01	-0.02	---
	Nationaliy: Remaining 'Muslim' countries	0.00	0.00	0.00	0.01	---
	Nationaliy: Remaining Asia	0.01	0.02*	0.01	0.01	---
	Nationaliy: Remaining countries	0.00	0.01	0.01	-0.01	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Information on cohort 6 in 2006/07 is missing since these students would have to be in grade 14 (which does not exist).
Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3.5: First-stage results – Population of students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.40**	0.42**	0.42**	0.41**	0.41**
(1998)	(F)	(1222)	(1305)	(1217)	(1149)	(1120)
	Observations	11090	10790	10850	10835	10630
Cohort 2	Coefficient	0.41**	0.41**	0.42**	0.42**	0.41**
(1997)	(F)	(1157)	(1082)	(1147)	(1113)	(1037)
	Observations	10335	10417	10480	10518	10192
Cohort 3	Coefficient	0.33**	0.33**	0.33**	0.33**	0.31**
(1996)	(F)	(798)	(780)	(692)	(667)	(597)
	Observations	10926	10947	11049	10905	10905
Cohort 4	Coefficient	0.31**	0.31**	0.31**	0.31**	0.32**
(1995)	(F)	(693)	(655)	(672)	(651)	(626)
	Observations	11064	11078	10788	10788	10788
Cohort 5	Coefficient	0.33**	0.34**	0.34**	0.34**	0.34**
(1994)	(F)	(849)	(820)	(771)	(762)	(685)
	Observations	10753	10400	10400	10400	10400
Cohort 6	Coefficient	0.34**	0.34**	0.34**	0.33**	---
(1993)	(F)	(784)	(755)	(724)	(685)	---
	Observations	10253	10253	10253	10253	---
		Results without lost observations				
Cohort 3	Coefficient	0.33**	0.33**	0.33**	0.33**	0.30**
(1996)	(F)	(798)	(780)	(692)	(667)	(499)
	Observations	10926	10947	11049	10905	9853
Cohort 4	Coefficient	0.31**	0.31**	0.31**	0.31**	0.33**
(1995)	(F)	(693)	(655)	(672)	(557)	(520)
	Observations	11064	11078	10788	10001	9345
Cohort 5	Coefficient	0.33**	0.34**	0.33**	0.33**	0.33**
(1994)	(F)	(849)	(820)	(714)	(560)	(450)
	Observations	10753	10400	10054	8872	8086
Cohort 6	Coefficient	0.34**	0.34**	0.35**	0.33**	---
(1993)	(F)	(784)	(658)	(624)	(471)	---
	Observations	10253	9464	8722	7812	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.6: OLS results – Population of students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	-0.02**	-0.09**	-0.11**	-0.12**	-0.12**
(1998)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	11090	10790	10850	10835	10630
Cohort 2	Coefficient	-0.11**	-0.12**	-0.12**	-0.13**	-0.12**
(1997)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	10335	10417	10480	10518	10192
Cohort 3	Coefficient	-0.09**	-0.11**	-0.11**	-0.11**	-0.04**
(1996)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	10926	10947	11049	10905	10905
Cohort 4	Coefficient	-0.11**	-0.11**	-0.11**	-0.04**	-0.04**
(1995)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	11064	11078	10788	10788	10788
Cohort 5	Coefficient	-0.10**	-0.08**	0.00	-0.04**	-0.09**
(1994)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	10753	10400	10400	10400	10400
Cohort 6	Coefficient	-0.08**	-0.02**	-0.04**	-0.07**	---
(1993)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	---
	Observations	10253	10253	10253	10253	---
Results without lost observations						
Cohort 3	Coefficient	-0.09**	-0.11**	-0.11**	-0.11**	-0.01
(1996)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	10926	10947	11049	10905	9853
Cohort 4	Coefficient	-0.11**	-0.11**	-0.11**	-0.02**	0.01
(1995)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	11064	11078	10788	10001	9345
Cohort 5	Coefficient	-0.10**	-0.08**	0.01	0.02**	-0.01*
(1994)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	10753	10400	10054	8872	8086
Cohort 6	Coefficient	-0.08**	0.00	0.04**	0.02*	---
(1993)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	---
	Observations	10253	9464	8722	7812	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.7: Second-stage results – Population of students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.13**	0.15**	0.16**	0.12**	0.11**
(1998)	(s.e.)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
	Observations	11090	10790	10850	10835	10630
Cohort 2	Coefficient	0.09**	0.10**	0.08**	0.08**	0.08**
(1997)	(s.e.)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
	Observations	10335	10417	10480	10518	10192
Cohort 3	Coefficient	0.13**	0.13**	0.12**	0.14**	0.05
(1996)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	10926	10947	11049	10905	10905
Cohort 4	Coefficient	0.19**	0.15**	0.14**	0.07**	0.05
(1995)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	0.03
	Observations	11064	11078	10788	10788	10788
Cohort 5	Coefficient	0.14**	0.14**	0.09**	0.03	0.00
(1994)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	0.03
	Observations	10753	10400	10400	10400	10400
Cohort 6	Coefficient	0.16**	0.03	0.00	0.02	---
(1993)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	---
	Observations	10253	10253	10253	10253	---
		Results without lost observations				
Cohort 3	Coefficient	0.13**	0.13**	0.12**	0.14**	0.11**
(1996)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
	Observations	10926	10947	11049	10905	9853
Cohort 4	Coefficient	0.19**	0.15**	0.14**	0.10**	0.08**
(1995)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
	Observations	11064	11078	10788	10001	9345
Cohort 5	Coefficient	0.14**	0.14**	0.11**	0.09**	0.05
(1994)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
	Observations	10753	10400	10054	8872	8086
Cohort 6	Coefficient	0.16**	0.05	0.09**	0.04	---
(1993)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.04)	---
	Observations	10253	9464	8722	7812	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.8: Second-stage results for change to grammar school as the outcome – Population of students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.13**	0.01	0.02	-0.01**	0.00
(1998)	(s.e.)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
	Observations	11077	10780	10842	10824	10630
Cohort 2	Coefficient	0.02**	0.00	0.00	0.00	0.00
(1997)	(s.e.)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
	Observations	10318	10412	10476	10512	10192
Cohort 3	Coefficient	0.00	0.02**	0.00	0.01	-0.08**
(1996)	(s.e.)	(0.01)	(0.01)	(0.00)	(0.00)	(0.02)
	Observations	10923	10934	11044	10902	10902
Cohort 4	Coefficient	0.01*	-0.01**	-0.01*	-0.04	-0.03*
(1995)	(s.e.)	(0.01)	(0.00)	(0.00)	(0.02)	(0.02)
	Observations	11061	11069	10787	10787	10787
Cohort 5	Coefficient	0.00	0.01*	-0.05**	-0.03*	-0.02*
(1994)	(s.e.)	(0.00)	(0.00)	(0.02)	(0.02)	(0.01)
	Observations	10744	10396	10396	10396	10396
Cohort 6	Coefficient	0.00	-0.06**	0.00	0.01	---
(1993)	(s.e.)	(0.00)	(0.02)	(0.02)	(0.01)	---
	Observations	10248	10248	10248	10248	---
		Results without lost observations				
Cohort 3	Coefficient	0.00	0.02**	0.00	0.01	-0.08**
(1996)	(s.e.)	(0.01)	(0.01)	(0.00)	(0.00)	(0.03)
	Observations	10923	10934	11044	10902	9853
Cohort 4	Coefficient	0.01*	-0.01**	-0.01*	-0.04	-0.02
(1995)	(s.e.)	(0.01)	(0.00)	(0.00)	(0.03)	(0.02)
	Observations	11061	11069	10787	10001	9345
Cohort 5	Coefficient	0.00	0.01*	-0.04*	-0.03	-0.02
(1994)	(s.e.)	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
	Observations	10744	10396	10053	8871	8086
Cohort 6	Coefficient	0.00	-0.06**	0.02	0.01	---
(1993)	(s.e.)	(0.00)	(0.02)	(0.02)	(0.02)	---
	Observations	10248	9463	8722	7812	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.9: Second-stage results for grammar school attendance as the outcome – Students sorted by grade attended – Population of students born in June or July

		2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.15**	0.14**	0.12**	0.16**
(1998)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11217	11513	11580	11022
Cohort 2	Coefficient	0.15**	0.12**	0.12**	0.10**
(1997)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11790	11832	11641	11215
Cohort 3	Coefficient	0.12**	0.13**	0.13**	0.04
(1996)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11565	11388	12098	13575
Cohort 4	Coefficient	0.13**	0.11**	0.05	0.02
(1995)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11136	12156	13564	13564
Cohort 5	Coefficient	0.12**	0.03	0.02	0.03
(1994)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11772	12909	12909	12909
Cohort 6	Coefficient	---	---	---	---
(1993)	(s.e.)	---	---	---	---
	Observations	---	---	---	---
Results without lost observations					
Cohort 3	Coefficient	0.12**	0.13**	0.13**	0.11**
(1996)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11565	11388	12098	12230
Cohort 4	Coefficient	0.13**	0.11**	0.04	0.04
(1995)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11136	12156	12481	12604
Cohort 5	Coefficient	0.12**	0.06*	0.05	0.03
(1994)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	11772	12187	10043	10496
Cohort 6	Coefficient	0.05	0.02	0.03	---
(1993)	(s.e.)	(0.03)	(0.03)	(0.04)	---
	Observations	11636	10241	7769	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 11 to 13 (see Section 3.3.3). The number of observations rise between grade 10 and 11 in the original data probably due to wrong information on the actual grade (which is why we do not use the grade information for the results presented in the remaining parts of this study). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations. No simulations are available for cohort 6 since we do not observe the required reference group of 10^{th} graders in 2002/03.
Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2003/04 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3.10: Second-stage results for narrow definition of grammar school attendance – Population of students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.13**	0.15**	0.16**	0.12**	0.11**
(1998)	(F)	0.02	0.02	0.03	0.03	(0.03)
	Observations	11090	10790	10850	10835	10630
Cohort 2	Coefficient	0.09**	0.10**	0.08**	0.08**	0.08**
(1997)	(F)	0.02	0.03	0.02	0.02	(0.03)
	Observations	10335	10417	10480	10518	10192
Cohort 3	Coefficient	0.13**	0.13**	0.12**	0.14**	0.11**
(1996)	(s.e.)	0.03	0.03	0.03	0.03	(0.03)
	Observations	10926	10947	11049	10905	10905
Cohort 4	Coefficient	0.19**	0.15**	0.14**	0.09**	0.08**
(1995)	(s.e.)	0.03	0.03	0.03	(0.03)	(0.03)
	Observations	11064	11078	10788	10788	10788
Cohort 5	Coefficient	0.14**	0.13**	0.11**	0.08**	0.07**
(1994)	(s.e.)	0.03	0.03	(0.03)	(0.03)	(0.03)
	Observations	10753	10400	10400	10400	10400
Cohort 6	Coefficient	0.16**	0.09**	0.06**	0.08**	---
(1993)	(s.e.)	0.03	(0.03)	(0.03)	(0.03)	---
	Observations	10253	10253	10253	10253	---
Results without lost observations						
Cohort 3	Coefficient	0.13**	0.13**	0.12**	0.14**	0.16**
(1996)	(F)	0.03	0.03	0.03	0.03	(0.04)
	Observations	10926	10947	11049	10905	9853
Cohort 4	Coefficient	0.19**	0.15**	0.14**	0.11**	0.11**
(1995)	(F)	0.03	0.03	0.03	0.04	(0.03)
	Observations	11064	11078	10788	10001	9345
Cohort 5	Coefficient	0.14**	0.13**	0.13**	0.13**	0.13**
(1994)	(F)	0.03	0.03	0.03	0.04	(0.04)
	Observations	10753	10400	10054	8872	8086
Cohort 6	Coefficient	0.16**	0.12**	0.13**	0.11**	---
(1993)	(F)	0.03	0.03	0.03	0.04	---
	Observations	10253	9464	8722	7812	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3.11: The return to a grammar school degree

	Men		Women	
Any Grammar School Degree	0.21**	0.19**	0.24**	0.25**
(s.e.)	(0.04)	(0.07)	(0.05)	(0.08)
Grammar no *Fachoberschule*	---	0.03	---	-0.01
(s.e.)	---	(0.08)	---	(0.09)
Age	0.15*	0.15*	0.07	0.07
(s.e.)	(0.08)	(0.08)	(0.10)	(0.10)
Age squared	-0.00	-0.00	-0.00	-0.00
(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.65	-0.65	1.07	1.06
(s.e.)	(1.31)	(1.32)	(1.61)	(1.60)
Observations	1462	1462	1172	1172
R^2	0.13	0.13	0.0592	0.0592

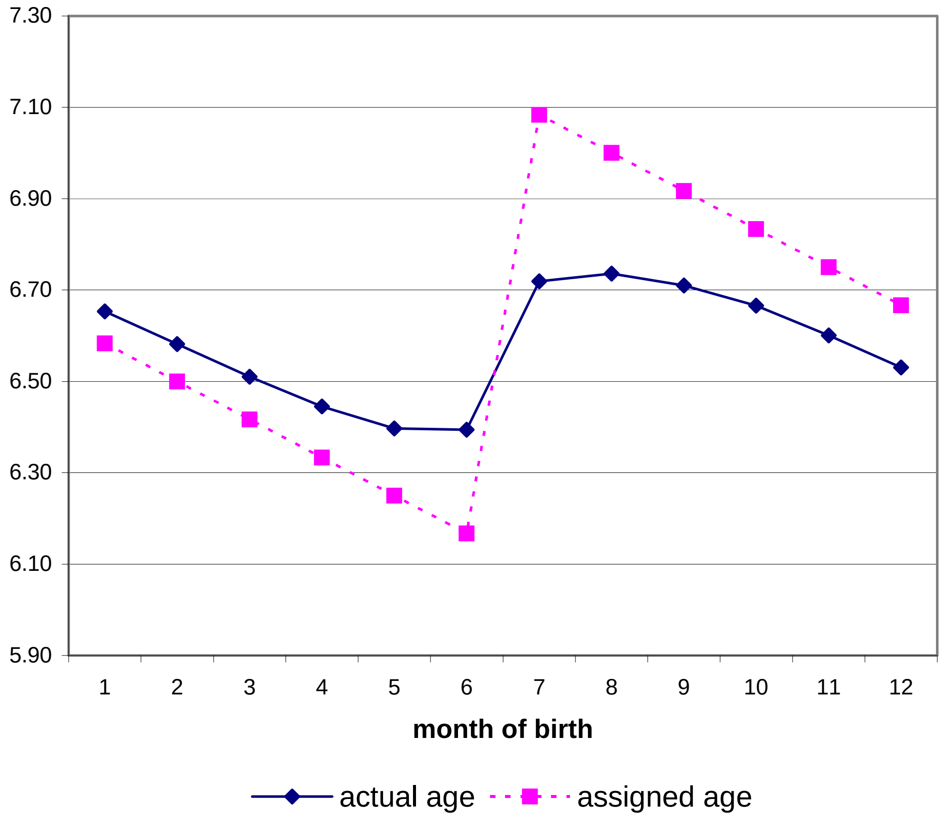
Note: The dependent variable is the natural logarithm of the gross hourly wage rate. Only West Germans aged 26 to 40 are included in the sample. *Any Grammar School Degree* is a dummy variable that equals 1 if a person holds a degree obtained from traditional *Gymnasium*, *berufliches Gymnasium*, or *Fachoberschule*. *Grammar no Fachoberschule* is a dummy variable that equals 1 if a person holds a degree obtained from traditional *Gymnasium* or *berufliches Gymnasium*. The German Socio-Economic Panel does not distinguish between degrees from traditional *Gymnasium* and *berufliches Gymnasium*. Estimates are obtained using sampling weights and robust standard errors.

* Significant at the ten percent level. ** Significant at the five percent level.

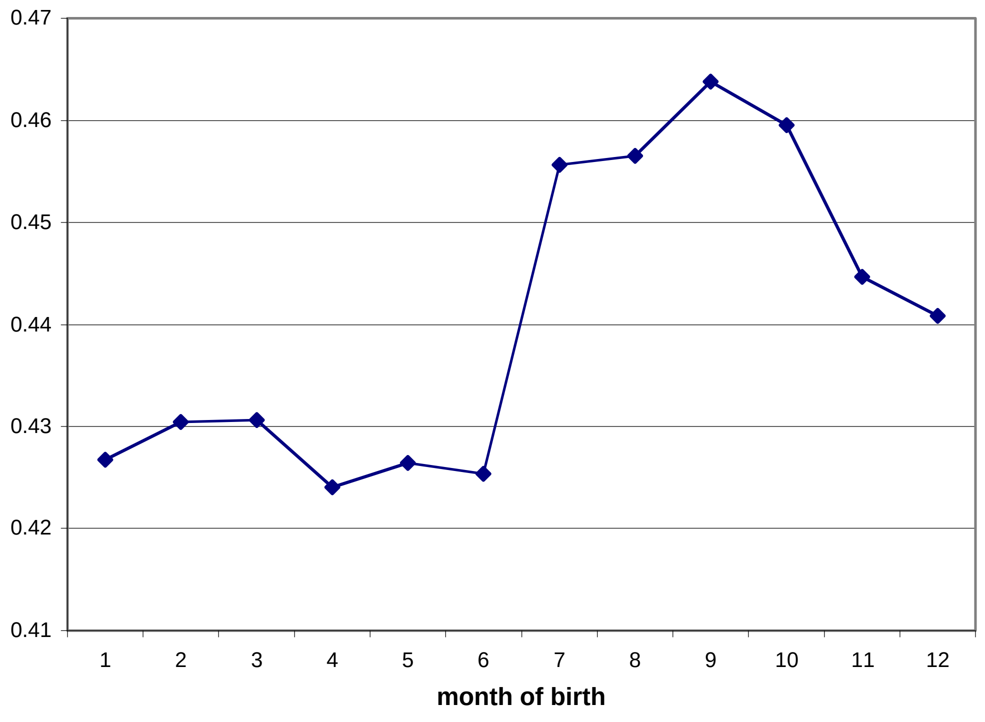
Source: German Socio-Economic Panel 2004. Own calculations.

Figure 3.1: The Hamburg Accord and educational outcomes

Actual and assigned school entry ages by month of birth

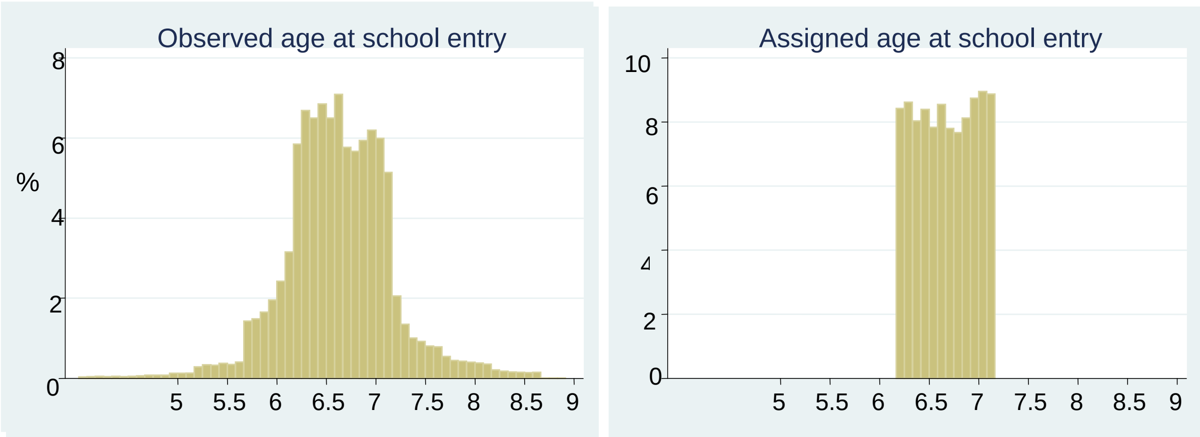


Grammar school attendance by month of birth



Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2005/06 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Figure 3.2: Observed and assigned age at school entry



Note: Assigned age at school entry according to the ‘Hamburg Accord’.
Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2005/06 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Appendix to Chapter 3

Table 3A.1: Simple correlations between instrument and observables – Full population

	Variable	2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1		0.00	0.00	0.00	0.00	0.00
(1998)	Male	0.00	0.00	0.00	0.00	0.00
	Region 1: (Darmstadt)	-0.01**	0.00	-0.01*	-0.01**	-0.01*
	Region 2: (Frankfurt)	0.00	0.00	0.01	0.01	0.00
	Region 3: (Offenbach)	0.00	0.00	0.00	0.00	0.00
	Region 4: (Wiesbaden, Main-Taun., Rheing.)	0.00	0.00	0.00	0.00	0.00
	Region 5: (Bergstr., Odenw., Diebg., G.-Ger.)	0.00	0.01	0.01*	0.01*	0.01
	Region 6: (Hochtaunus, Wetterau)	0.00	0.00	0.00	0.00	0.00
	Region 7: (Main-Kinzig)	-0.01	0.00	0.00	0.00	0.00
	Region 8: (Gießen, Lahn-Dill, Limburg-Weil.)	0.00	0.00	0.00	0.00	0.00
	Region 9: (Marburg-Biedenkopf, Vogelsberg)	0.00	0.00	0.00	0.00	0.01
	Region 10: (Kassel)	0.00	0.00	0.00	0.00	0.00
	Region 11: (Fulda, Hersfeld-Rotenburg)	0.00	-0.01	-0.01	-0.01	0.00
	Region 12: (Kassel-Land, Werra-Mei., Sch.-E)	0.01**	0.01**	0.01**	0.01**	0.02**
	Nationaliy: German speaking country	-0.02**	-0.02**	-0.01**	-0.02**	-0.02**
	Nationaliy: Turkey	-0.01	0.00	-0.01	-0.01	-0.01
	Nationaliy: Italy, Greece	0.00	0.00	0.00	0.01	0.01*
	Nationaliy: Former Yugoslavian states	0.00	0.01	0.00	0.00	0.00
	Nationaliy: Remaining 'Western' countries	0.00	0.00	0.00	0.00	-0.01
	Nationaliy: Eastern Europe, former Soviet Un.	0.00	0.00	-0.01	0.00	0.00
	Nationaliy: Remaining 'Muslim' countries	0.00	0.00	0.01	0.00	0.00
	Nationaliy: Remaining Asia	0.00	0.00	0.00	-0.01	0.00
	Nationaliy: Remaining countries	-0.01	-0.01	0.00	0.00	---
Cohort 6		0.00	0.00	-0.01**	-0.01*	---
(1993)	Male	0.01*	0.01**	0.01	0.00	---
	Region 1: (Darmstadt)	0.00	-0.01*	0.00	0.01	---
	Region 2: (Frankfurt)	0.00	0.00	0.00	0.00	---
	Region 3: (Offenbach)	0.00	0.00	0.00	0.00	---
	Region 4: (Wiesbaden, Main-Taun., Rheing.)	0.01	0.01**	0.01	0.00	---
	Region 5: (Bergstr., Odenw., Diebg., G.-Ger.)	0.00	0.00	-0.01*	-0.01*	---
	Region 6: (Hochtaunus, Wetterau)	-0.01*	0.00	0.00	0.00	---
	Region 7: (Main-Kinzig)	0.01**	0.01**	0.01	0.00	---
	Region 8: (Gießen, Lahn-Dill, Limburg-Weil.)	0.00	0.00	0.00	0.00	---
	Region 9: (Marburg-Biedenkopf, Vogelsberg)	-0.01	-0.01*	0.00	0.01	---
	Region 10: (Kassel)	0.00	0.00	0.00	0.00	---
	Region 11: (Fulda, Hersfeld-Rotenburg)	0.01**	0.01	0.01	0.00	---
	Region 12: (Kassel-Land, Werra-Mei., Sch.-E)	-0.01*	-0.01*	-0.01*	-0.01	---
	Nationaliy: German speaking country	0.00	0.01**	0.00	0.00	---
	Nationaliy: Turkey	0.00	0.00	0.00	0.00	---
	Nationaliy: Italy, Greece	0.00	-0.01	0.00	0.00	---
	Nationaliy: Former Yugoslavian states	0.00	-0.01	0.00	0.00	---
	Nationaliy: Remaining 'Western' countries	-0.01	0.00	0.00	0.00	---
	Nationaliy: Eastern Europe, former Soviet Un.	0.00	0.00	0.01	0.00	---
	Nationaliy: Remaining 'Muslim' countries	0.00	0.00	0.00	0.00	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Information on cohort 6 in 2006/07 is missing since these students would have to be in grade 14 (which does not exist).
Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3A.2: First-stage results – Full population of students

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.44**	0.48**	0.48**	0.47**	0.44**
(1998)	(F)	(3904)	(4391)	(4173)	(3958)	(3404)
	Observations	62832	61438	61552	61398	61947
Cohort 2	Coefficient	0.49**	0.49**	0.48**	0.48**	0.46**
(1997)	(F)	(4552)	(4190)	(4095)	(3911)	(3509)
	Observations	59194	59834	60114	60917	59628
Cohort 3	Coefficient	0.32**	0.32**	0.33**	0.33**	0.31**
(1996)	(F)	(1858)	(1889)	(1790)	(1707)	(1525)
	Observations	63425	63621	63937	63240	63240
Cohort 4	Coefficient	0.32**	0.32**	0.33**	0.33**	0.34**
(1995)	(F)	(1918)	(1870)	(1970)	(1881)	(1795)
	Observations	64037	64003	62735	62735	62735
Cohort 5	Coefficient	0.34**	0.34**	0.35**	0.34**	0.33**
(1994)	(F)	(2210)	(2167)	(2096)	(1946)	(1746)
	Observations	62673	60941	60941	60941	60941
Cohort 6	Coefficient	0.34**	0.35**	0.35**	0.33**	---
(1993)	(F)	(1961)	(1987)	(1915)	(1734)	---
	Observations	58599	58599	58599	58599	---
		Results without lost observations				
Cohort 3	Coefficient	0.32**	0.32**	0.33**	0.33**	0.31**
(1996)	(F)	(1858)	(1889)	(1790)	(1707)	(1363)
	Observations	63425	63621	63937	63240	57890
Cohort 4	Coefficient	0.32**	0.32**	0.33**	0.32**	0.34**
(1995)	(F)	(1918)	(1870)	(1970)	(1612)	(1500)
	Observations	64037	64003	62735	58557	54669
Cohort 5	Coefficient	0.34**	0.34**	0.33**	0.35**	0.34**
(1994)	(F)	(2210)	(2167)	(1812)	(1696)	(1285)
	Observations	62673	60941	58700	51945	47305
Cohort 6	Coefficient	0.34**	0.34**	0.38**	0.37**	---
(1993)	(F)	(1961)	(1723)	(1925)	(1472)	---
	Observations	58599	54887	51390	44653	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3A.3: OLS results – Full population of students

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	-0.01	-0.11**	-0.13**	-0.13**	-0.14**
(1998)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	62832	61438	61552	61398	61947
Cohort 2	Coefficient	-0.12**	-0.12**	-0.13**	-0.13**	-0.11**
(1997)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	59194	59834	60114	60917	59628
Cohort 3	Coefficient	-0.11**	-0.14**	-0.14**	-0.12**	-0.00
(1996)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	63425	63621	63937	63240	63240
Cohort 4	Coefficient	-0.14**	-0.15**	-0.14**	-0.02**	-0.00
(1995)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	64037	64003	62735	62735	62735
Cohort 5	Coefficient	-0.12**	-0.10**	0.00	-0.02**	-0.05**
(1994)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	62673	60941	60941	60941	60941
Cohort 6	Coefficient	-0.10**	-0.01*	-0.02**	-0.05**	---
(1993)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	---
	Observations	58599	58599	58599	58599	---
			Results without lost observations			
Cohort 3	Coefficient	-0.11**	-0.14**	-0.14**	-0.12**	0.02
(1996)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	63425	63621	63937	63240	57890
Cohort 4	Coefficient	-0.14**	-0.15**	-0.14**	-0.02**	0.02
(1995)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	64037	64003	62735	58557	54669
Cohort 5	Coefficient	-0.12**	-0.10**	0.01*	0.02**	0.00*
(1994)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Observations	62673	60941	58700	51945	47305
Cohort 6	Coefficient	-0.10**	0.00	0.02**	0.01	---
(1993)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	---
	Observations	58599	54887	51390	44653	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3A.4: Second-stage results – Full population of students

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.15**	0.19**	0.19**	0.18**	0.20**
(1998)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
	Observations	62832	61438	61552.00	61398	61947
Cohort 2	Coefficient	0.09**	0.08**	0.08**	0.08**	0.09**
(1997)	(s.e.)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Observations	59194	59834	60114	60917	59628
Cohort 3	Coefficient	0.19**	0.20**	0.18**	0.18**	0.08**
(1996)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	63425	63621	63937	63240	63240
Cohort 4	Coefficient	0.23**	0.21**	0.17**	0.08**	0.02
(1995)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	64037	64003	62735	62735	62735
Cohort 5	Coefficient	0.17**	0.15**	0.09**	0.03	0.01
(1994)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	62673	60941	60941	60941	60941
Cohort 6	Coefficient	0.22**	0.09**	-0.02**	-0.05**	---
(1993)	(s.e.)	(0.02)	(0.02)	(0.00)	(0.00)	---
	Observations	58599	58599	58599	58599	---
		Results without lost observations				
Cohort 3	Coefficient	0.19**	0.20**	0.18**	0.18**	0.13**
(1996)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	63425	63621	63937	63240	57890
Cohort 4	Coefficient	0.23**	0.21**	0.17**	0.09**	0.11**
(1995)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	64037	64003	62735	58557	54669
Cohort 5	Coefficient	0.17**	0.15**	0.09**	0.08**	0.05**
(1994)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Observations	62673	60941	58700	51945	47305
Cohort 6	Coefficient	0.22**	0.09**	0.12**	0.06**	---
(1993)	(s.e.)	(0.02)	(0.02)	(0.02)	(0.02)	---
	Observations	58599	54887	51390	44653	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3A.5: Second-stage results for change to grammar school – Full population of students

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.15**	0.03**	0.01	0.00	0.00
(1998)	(s.e.)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
	Observations	62757	61368	61505	61347	61947
Cohort 2	Coefficient	0.04**	0.00	0.01**	0.00	0.00*
(1997)	(s.e.)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
	Observations	59125	59781	60087	60882	59628
Cohort 3	Coefficient	-0.01	0.04**	0.00*	0.00	-0.11**
(1996)	(s.e.)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)
	Observations	63391	63577	63913	63224	63224
Cohort 4	Coefficient	0.04**	0.00	0.00*	-0.07**	-0.02*
(1995)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
	Observations	63992	63960	62731	62731	62731
Cohort 5	Coefficient	0.00	0.00	-0.07**	-0.02**	-0.02**
(1994)	(s.e.)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
	Observations	62641	60913	60913	60913	60913
Cohort 6	Coefficient	0.00	-0.08**	-0.01	-0.01	---
(1993)	(s.e.)	(0.00)	(0.01)	(0.01)	(0.01)	---
	Observations	58569	58569	58569	58569	---
		Results without lost observations				
Cohort 3	Coefficient	-0.01	0.04**	0.00*	0.00	-0.11**
(1996)	(s.e.)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)
	Observations	63391	63577	63913	63224	57890
Cohort 4	Coefficient	0.04**	0.00	0.00*	-0.08**	-0.01
(1995)	(s.e.)	(0.00)	(0.00)	(0.00)	(0.02)	(0.01)
	Observations	63992	63960	62731	58552	54669
Cohort 5	Coefficient	0.00	0.00	-0.08**	-0.02*	-0.02**
(1994)	(s.e.)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)
	Observations	62641	60913	58693	51939	47305
Cohort 6	Coefficient	0.00	-0.10**	0.00	-0.01	---
(1993)	(s.e.)	(0.00)	(0.02)	(0.01)	(0.01)	---
	Observations	58569	54884	51390	44653	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (Hessisches Statistisches Landesamt). Own calculations.

Table 3A.6: Second-stage results – Population of male students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.15**	0.14**	0.14**	0.11**	0.10**
(1998)	(s.e.)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
	Observations	5553	5369	5407	5411	5279
Cohort 2	Coefficient	0.08**	0.09**	0.08**	0.07**	0.07**
(1997)	(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Observations	5229	5282	5299	5329	5092
Cohort 3	Coefficient	0.16**	0.14**	0.13**	0.17**	0.10*
(1996)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
	Observations	5586	5587	5634	5528	5528
Cohort 4	Coefficient	0.17**	0.13**	0.14**	0.06	0.00
(1995)	(s.e.)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
	Observations	5579	5555	5396	5396	5396
Cohort 5	Coefficient	0.14**	0.14**	0.09*	0.02	0.05
(1994)	(s.e.)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
	Observations	5492	5226	5226	5226	5226
Cohort 6	Coefficient	0.13**	0.00	0.00	0.04	---
(1993)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	---
	Observations	5252	5252	5252	5252	---
		Results without lost observations				
Cohort 3	Coefficient	0.16**	0.14**	0.13**	0.17**	0.14**
(1996)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
	Observations	5586	5587	5634	5528	4984
Cohort 4	Coefficient	0.17**	0.13**	0.14**	0.08	0.06
(1995)	(s.e.)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
	Observations	5579	5555	5396	4993	4653
Cohort 5	Coefficient	0.14**	0.14**	0.12**	0.09*	0.07
(1994)	(s.e.)	(0,04)	(0,04)	(0,05)	(0,05)	(0,05)
	Observations	5492	5226	5159	4553	4160
Cohort 6	Coefficient	0,13**	0,05	0,10**	0,06	---
(1993)	(s.e.)	(0,04)	(0,04)	(0,05)	(0,05)	---
	Observations	5252	4903	4460	4025	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (see Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they sought a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Table 3A.7: Second-stage results – Population of female students born in June or July

		2002/03	2003/04	2004/05	2005/06	2006/2007
Cohort 1	Coefficient	0.11**	0.16**	0.18**	0.18**	0.13**
(1998)	(s.e.)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
	Observations	5537	5421	5443	5443	5351
Cohort 2	Coefficient	0.11**	0.09**	0.08**	0.08**	0.08**
(1997)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	Observations	5106	5135	5181	5181	5100
Cohort 3	Coefficient	0.11**	0.12**	0.11**	0.11**	0.00
(1996)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
	Observations	5340	5360	5415	5415	5377
Cohort 4	Coefficient	0.20**	0.16**	0.14**	0.07	0.07*
(1995)	(s.e.)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
	Observations	5485	5532	5392	5392	5392
Cohort 5	Coefficient	0.13**	0.13**	0.07	0.01	-0.06
(1994)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	Observations	5261	5174	5174	5174	5174
Cohort 6	Coefficient	0.18**	0.04	-0.06	-0.01	---
(1993)	(s.e.)	(0.05)	(0.05)	(0.05)	(0.05)	---
	Observations	5001	5001	5001	5001	---
		Results without lost observations				
Cohort 3	Coefficient	0.11**	0.12**	0.11**	0.11**	0.07
(1996)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
	Observations	5340	5360	5415	5415	4869
Cohort 4	Coefficient	0.20**	0.16**	0.14**	0.10**	0.08*
(1995)	(s.e.)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
	Observations	5485	5532	5392	5008	4692
Cohort 5	Coefficient	0.13**	0.13**	0.08*	0.06	0.03
(1994)	(s.e.)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
	Observations	5261	5174	4895	4319	3926
Cohort 6	Coefficient	0.18**	0.03	0.07	0.01	---
(1993)	(s.e.)	(0.05)	(0.05)	(0.05)	(0.05)	---
	Observations	5001	4561	4262	3787	---

Note: * Significant at the ten percent level. ** Significant at the five percent level. Documented coefficients refer to specifications without control variables. Effects are robust if available control variables (gender, region and nationality) are considered. The upper panel of the estimates includes simulation results holding the number of observations constant for grades 10 to 13 (compare Section 3.3.3). Missing observations are assumed to be lower track students, since they would be in the data if they seeked a higher secondary degree. The lower panel of the estimates shows the results based only on observed students without simulated observations.

Source: Student-Level Data of the Statistics of General Schools for the State of *Hessen* 2002/03 to 2006/07 provided by the State Statistical Office (*Hessisches Statistisches Landesamt*). Own calculations.

Chapter 4

An Evaluation of Single and Mixed Gender Computer Science Classes *

* This chapter is based on an earlier discussion paper version: Discussion Paper of the Department of Economics, Leibniz Universität Hannover, 369, 2007.

Abstract: Discussions on the benefits of single-gender education on girls' science outcomes are popular in the German education literature. However, most empirical evidence tends to be qualitative work and the causal effects of single-gender education are hardly identified using appropriate statistical methods. This chapter provides insights from a recent single-gender-education school project conducted in computer science classes at a German lower secondary school. About 80 students participated in this intervention study repeatedly answering specifically designed questionnaires and tests. The project fails to identify positive effects from single-gender education but the interpretation is impeded by several confounding factors. When directly asked, most students prefer to be educated in mixed-gender groups, while the participating teachers judge their teaching experience with the project groups in favour of single-gender education.

JEL classification: I21, J16

Keywords: gender, education, identification, coeducation, segregation, experiments

Acknowledgement: This research was partly supported by the Anglo-German Foundation within the project 'The Economics and Politics of Employment, Migration and Social Justice', which is part of the Foundation's research initiative "Creating Sustainable Growth in Europe". I am obliged to the school principal, teachers, pupils, parents, and those responsible at the local school authority who kindly supported this project. I am grateful to Patrick Puhani for encouraging the study and for valuable comments. I also thank Horst Entorf, Olaf Hübler, Christian Pfeifer and participants of the Ph.D. Seminars in Applied Labor Economics at Darmstadt University of Technology and at the Leibniz University of Hannover for helpful comments as well as Björn Schumacher and Philip Savage for excellent research assistance. All remaining errors in this paper are my own.

Table of Contents for Chapter 4

4.1 Motivation of the Intervention Study and Stylized Facts

Coeducation was broadly introduced in West Germany in the 1950s and 1960s and in East Germany in 1945 and was taken to be an important measure for the assurance of equal educational and professional opportunities for both genders. However, educational experts soon started to doubt the universal benefits of mixed-gender education. In the 1980s, German universities realised that the overwhelming proportion of girls choosing technical study programmes had graduated from single-gender high schools. In 1989, the German feminist magazine "EMMA" even came up with the slogan 'coeducation makes girls stupid' (*Koedukation macht Mädchen dumm*)[78]. In the same year, the federal states' women's representatives (*Frauenbeauftragte der Länder*) suggested that instruction techniques should eliminate gender stereotypes in the classroom and compensate disadvantages for female students (cf. Volmerg *et al*., 1996, p.11).

Thus, debates on the advantages and disadvantages of coeducation and how to design coeducation made a return to the political agenda in the 1980s. These discussions were accompanied by evolving educational, psychological and sociological research studies on coeducation and educational performance by gender. The classic arguments used in this education literature to explain girls' reluctance to choose technical subjects are (1) little experience related to the contents of these subjects before entering, or outside of, the classroom combined with the feeling that they are unable to catch up with the boys' head start[79], (2) a self-critical assessment of their own abilities, which is typically observed for female students[80], (3) a lack of *female* role models related to the respective subjects and (4) a lack of educational techniques focusing on the needs or interests of female students.

Theoretically, there are two major ways out of this coeducational dilemma: First of all, researchers suggested going back to single-gender education in technical subjects. In fact, nowadays (beside the principle of general coeducation of the genders) the different states' education laws (*Schulgesetze*) explicitly include the possibility of (temporary) single-gender education in some subjects. The second option is to foster instruction techniques *within* co-

---

[78] Quoted according to Kraus (1998). A comparable phrase that spread in English-speaking countries is 'better dead than coed'.

[79] Compare for example Schuld (1997) and Heidtmann (1998) for gender differences in computer use at home. From this literature it is obvious that teenage boys more often possess a computer than girls and use computers in their leisure time.

[80] If students are asked to judge their own performance, girls generally judge their performance worse compared to boys' self-assessment (given the same state of knowledge). Generally, there is also a consensus in the education literature that boys and girls behave differently in class: Girls are more often cooperative and behave according to teachers' expectations, while boys show more competitiveness. Compare for example Funken *et al.* (1996), Volmerg *et al.* (1996) or Rustmeyer und Jubel (1996) for discussions of these phenomena.

education classes which focus on the specific needs and interests of girls. The present paper follows the first approach examining whether coeducation really provides means to improve girls' performance in typically male-dominated subjects.

Until now, research on the benefits of single-gender education mainly focused on higher level secondary schools (*Gymnasium*). However, it is a fact that the group of graduates from the lower level schools (*Hauptschule*) experiences most problems in the labour market (*e.g.* high unemployment and low labour earnings). Girls (specifically those with an immigrant background) are known to be an especially disadvantaged group among lower level secondary school students: They are often raised according to traditional gender role models and are characterized by low social mobility (cf. Brendel, 1995 and Thierack, 1995). Therefore, it is crucial to develop specific educational measures in order to improve the educational situation of these 'losers' in the education system.[81] Single-gender education might be one promising measure in order to improve the situation of poorly educated girls.

Table 4.1: Popular apprenticeship choices among lower secondary school graduates

Male		Female	
Subject of apprenticeship	%	Subject of apprenticeship	%
Vehicle Construction and Maintenance	12.75	Health and Hygiene	32.20
Construction	9.11	Retail	17.71
Painters, Varnishers	7.79	Commerce	13.46
Metal Processing	7.37	Hotel and Catering Industry	8.18
Mining, Manufacturing	6.98	Clerical Employees	8.08
Plumbers, Installers	6.71	Agriculture, Farming, Forestry	4.24
Electronics	6.29	Textile, Clothing, Nutrition (without cooks)	3.03
Commerce	5.87	Cooks	2.45
Textile, Clothing, Nutrition (without cooks)	5.56	Transportation, Stocking, Cleaning	1.42
Transportation, Stocking, Cleaning	5.43	Mining, Manufacturing	1.01

Note: The figures refer to the south-western German state of *Hessen* for which the required data is available. There are missing observations on the subject of apprenticeships for 2.66 % of the male students and 4.06 % of the female students.
Source: Student-Level Data of the Statistics of General Schools of the State of *Hessen* 2005/2006 provided by the State Statistical Office. Sample of 24,383 male and 11,116 female lower secondary school graduates in vocational schools. Own calculations.

Table 4.1 presents evidence on the ten most popular apprenticeship professions of graduates of the lower secondary schools in one German state by gender.[82] Obviously, girls' apprenticeship choices are much more focused on specific professional fields than boys' choices. Two-thirds of the female lower secondary graduates work in the fields of health and hygiene, retail

[81] The lack of studies examining lower secondary school students has been discussed before, for example in the papers by Brendel (1995) and Thierack (1995).
[82] The numbers refer to apprentices in the German state of *Hessen* since this is the only state for which the required information on apprenticeship choice and attained secondary degree is available from the official school statistics. Aggregated statistics for all apprentices independent of the secondary degree they hold are provided by the Federal Statistical Office and yield a similar pattern. Examination of industry codes of former lower secondary school graduates in the German-Socio Economic Panel (GSOEP) also yields similar results.

or commerce, where they are trained for 'classical' female professions like nurses, saleswomen or secretaries. Male students on the other hand tend to be trained in more technical professions (*e.g.* vehicle construction) which are hardly on the list of female graduates' choices. Additionally, male graduates of the lower secondary schools spread to a larger variety of apprenticeships compared to the female graduates. All in all, the choice of professions among lower secondary school graduates seems to be rather gender-specific and corresponds to classical gender-stereotypes.

In order to improve the labour market perspectives of female lower secondary school graduates it might be a promising step to foster their interest in the neglected study fields of technical subjects. As an example, and because of the growing importance of computer related skills in today's labour markets, the present paper focuses on computer studies (cf. for example Black and Spitz-Oener, 2007 for the change in gender-related job tasks in light of the technological progress). Specifically, the research question to be answered is whether girls' interest and performance in computer education may be improved through single-gender education. Therefore, a new school project was conducted during the school years 2004/2005 and 2005/2006 in an intermediate size lower level secondary school in the German state of **B**aden-**W**ürttemberg (hence the project name 'BW-project' which is used for the remainder of this paper). The intervention refers to computer science lessons of approximately eighty students who were in fifth grade in the first year of the project.

This paper proceeds as follows: First of all it reviews empirical strategies and existing evidence in order to assess the benefits and detriments of coeducation on girls' performance in computer studies (Section 4.2). The literature review focuses on German studies but also provides a summary of international evidence. Furthermore, the set-up of the BW-project is presented in Section 4.3 together with the findings of this project. When directly asked, most students prefer to be educated in mixed-gender groups, while the participating teachers judge their teaching experience with the project groups in favour of single-gender education. Considering different performance measures, the project does not allow conclusion in favour of coeducation. However, this interpretation is impeded by several confounding factors. Section 4.4 summarizes and discusses the current state of knowledge. The need for further empirical evidence is emphasized.

4.2 Review of Empirical Designs and Literature

The research question of interest is whether pupils, and specifically girls, benefit from single-gender science education. Generally, it is possible to measure the 'benefits' by focusing on different types of performance measures, which might be categorized as 'hard' and 'soft' measures[83]: Hard performance measures directly assess educational performance based, for example, on grades, test scores or students' self-assessment, while soft measures relate to behavioural or social outcomes like self-confidence, interest or motivation.

Given the set of outcome variables of interest, the empirical researcher aims to compare these outcomes of persons participating in specific single-gender education measures (the treated group) and persons not participating (the untreated) in order to identify a 'treatment effect'. For example, the outcomes (*e.g.* test scores, motivational or behavioural indicators) of students in single-gender classes are to be compared to outcomes of students in coeducational classes. Logically, it is not possible to observe identical students under both the single-gender regime and the coeducational regime at the same time. If one simply compares students from single-gender classes to students in mixed-gender classes, it is very likely that these students differ in observed or unobserved characteristics which may in turn influence the outcome variables. In other words: The observed effects will be biased because of the selection of students to the different groups. This is the typical evaluation problem which we have in the social sciences (compare for example the formalization in Heckman and Smith, 1995). Thus, in order to evaluate the impact of single-gender education, the crux is to find an adequate control group which did not receive single-gender treatment and does not distinguish from the treatment group in background characteristics determining educational success.

4.2.1 General Evaluation Strategies

How can the present evaluation problem thus be solved? Generally, there are three methodological bags of tricks differing in the way in which an adequate control group is created. The three strategies, which will be discussed for the present purpose, are: (1) *social experiments*, (2) *natural experiments,* and (3) *non-experimental approaches:*

Social experiments imply an intervention study where students are randomly assigned to coeducational and single-gender classes. Due to this procedure, it is expected that the

[83] Compare for example Hoffmann *et al.* (1997), chapter 6 for these different measures and their assessment in the German literature.

groups of treated students (single-gender education) and untreated students (mixed-gender education) do not differ in the distribution of their observed and unobserved characteristics.

A general critique with respect to the required randomisation refers to the fact that this procedure might change the pool of persons who are willing to participate in the evaluation study and the behaviour of participants. This effect is known as the *randomization bias* of social experiments. The major practical problem of randomisation is, however, that random assignment of students to classes may be disapproved by school principals or other participating persons: Sometimes random treatment (as compared to free choice of received treatment) is considered to be politically incorrect if there are concerns that the untreated students are arbitrarily deprived of beneficial measures. Furthermore, randomisation of students into groups (*e.g.* in the technical subjects considered) may be difficult for administrative reasons. Specifically, students are usually instructed in the same class context for all subjects in Germany which complicates the random assignment to different groups in a specific subject. Additionally, schools face limited resources in relation to, for example, teachers' work hours and the available science or computer classroom capacities.

As an alternative identification strategy, one stream of the existing literature considers ***natural experiments***. A natural experiment consists of an exogenous policy change affecting only subgroups of the population where assignment to the affected groups is exogenous to the outcome. Few previous studies drawing on natural experiments consider the effects that arise from a transition of schools from a coeducational to a single-gender regime or *vice versa*, as will be summarized in Section 4.2.3. Besides simply looking at changes over time, natural experiments can make use of pre-post designs, where performance changes of students experiencing a school's transformation (treatment group) are compared to the outcome patterns of students in similar schools or classes which did not experience such a regime transformation (control group). Thus, the chosen approach corresponds to a difference-in-difference approach where the performance change over time (first difference) is compared between a treatment and a comparison group (second difference). The lack of evidence related to natural experiments in the existing literature is symptomatic of the absence of such present regime changes.

Since natural experiments are rare and social experiments are hardly feasible, practical alternatives consist in ***non-experimental methods.*** Such studies draw on a comparison of students educated in single-gender and mixed-gender groups, explicitly taking initial selection into account via appropriate statistical methods. For example, matching techniques might be used in order to compare the outcomes of persons in treatment and control groups which are similar in (observed) individual background variables.[84] Regression adjustment methods regress the outcome of interest on a treatment dummy and individual background variables. Specifically, one might think of directly modelling initial selection into the treatment group *via* appropriate econometric techniques (*e.g.* using switching regressions, cf. Section 4.2.3).

4.2.2 General Implementation Issues

The detailed design of the intervention project needs to address different potential sources of biases. Such biases may occur independent of the underlying research strategy (social experiment, natural experiment or non-experimental evaluation). To start with, problems may emerge due to ***treatment substitution***, *i.e.* if the persons in the control group receive a substitutive treatment similar to the treatment under consideration.[85] A substitution bias would occur if, for example, students in single-gender (treatment) groups were taught by standard educational methods while students in mixed-gender (control) groups were taught by instruction methods focusing on girls' interests. In this case, the specific education methods are a substitute for the single-gender treatment and it is not possible to identify the treatment effect by comparing both groups. Generally***, multiple treatments*** may make it impossible to disentangle the effects of these different treatment measures. In order to avoid a substitution or multiple treatment bias, ideally all students have to be instructed in the same (or at least similar) way by the same (similar) teacher under the same (similar) circumstances.

The determination of treatment effects might also be complicated if the participating persons know that they are subjects of an evaluation study. Specifically, if students know that their results will be evaluated and related to their gender, this is likely to change their behaviour where the respective changes might differ by gender. As a consequence, (again) observed effects might be biased. This issue is referred to as an ***observation bias***. If students adjust their behaviour under treatment because they think this is what their *teachers* expect, this is called a ***Pygmalion effect*** in the education literature (cf. for example Ziegler *et al.*, 1998 for the discussion of the importance of this effect). In order to prevent these sources of biases, one possibility is to conceal the subject of observation and expected outcomes from the project participants.

A general problem of intervention studies considering a longer time span is that there might be significant sample attrition. Attrition will generate ***attrition biases*** in estimating the desired effects if the students dropping out of the sample systematically differ from the re-

[84] See, for example, Heckman *et al.* (1998) for a theoretical discussion of matching as an econometric evaluation estimator.

[85] See Heckman and Smith (1995) for a discussion of the substitution bias and other biases in social experiments.

maining students. Attrition may occur due, for example, to non-response to the research questions or if students who have initially been assigned to one of the compared groups are not satisfied with this educational situation and decide to leave the group. In the most extreme case students would leave the school or have to leave the school, respectively. Generally, it might be that school dropouts change the composition of the groups considered of interest.

4.2.3 Literature Review with a Focus on German Studies

A great many existing studies conducted by researchers from different social science disciplines around the world try to empirically evaluate the effects of single-gender education. Mael *et al.* (2005) provide a comprehensive literature and methods overview for industrialized countries. However, this survey article demonstrates that an overwhelming proportion of the present literature simply compares outcomes of students in single-gender and coeducational classes without taking selection to different educational regimes into account. In many cases the compared groups are even located at different schools. Thus, such simple correlation studies are likely to suffer from severe selection biases (cf. Section 4.2.1) and are not adequate for identifying a causal effect of single-gender education. One consequence of lacking identification strategies is that previous studies provide mixed evidence with some equivocal support of positive impacts of single-gender education related to school performance.

Since the survey by Mael *et al. (*2005) only considers evidence from studies in the English language and since the existing studies for Germany are mostly German-language research reports, hardly any evidence on Germany is summarized. Table 4E.1 in the ‘Appendix to the Literature Review’ provides an overview of recent research on single-gender computer education in Germany.[86] Due to the limited number of such studies, related fields of education (*i.e.* mathematics and science) are included in this review. Studies are categorized according to the way in which they solve the evaluation problem. The four categories considered are (1) correlation studies which do not solve the evaluation problem, (2) non-experimental evaluation studies drawing on matching or similar techniques, (3) natural experiments implying exogenous policy changes and (4) social experiments where students are randomly assigned to treatment and control groups.

The overview given in Table 4E.1 does not claim to be exhaustive: Especially, a large variety of articles (including results only published in newspaper articles) related to the ‘cor-

[86] While the focus of the following literature review is on the effects of coeducation on girls’ outcomes, fewer studies consider the impacts on male students. Holz-Ebeling *et al.* (2000) is an exceptional study focusing on boys’ outcomes. One conclusion of the study is that boys’ educational and social outcomes are hardly affected by single-gender education. However, since the paper examines boys in coeducational and single-gender educational schools without appropriately accounting for the selectivity into different school types, the findings are again to be interpreted with reservation.

relation study' category exists. However, because of the inherent problematic identification technique these studies are not of primary interest to the present paper. Additionally, most of the summarized papers do not solely concentrate on the specific topic related to the evaluation of coeducation *vs.* single-gender education but also cover more general topics (*e.g.* discussions of curricula or overviews of the history of coeducation). Since these special topics are not the focus of this paper, the given summary is restricted to relevant evaluation results. Moreover, only *recent* evidence, starting in the mid 1990s, is considered. One older study is presented when discussing the evidence from natural experiments since this is the newest available study for this category.[87] Finally, further recent German language publications refer to intervention studies in Austria or Switzerland (*e.g.* Faulstich-Wieland, 2004a provides recent evidence). Since the interest of this summary is related to studies for Germany, I do not summarize these papers.

The studies summarized in category (1) in Table 4E.1 are mostly qualitative reports including simple descriptive statistics. Funken *et al.* (1996), Meyer (1996) and Volmerg *et al.* (1996) explicitly focus on computer studies while Nyssen (1996) considers the related context of science classes. The size of the studies in terms of underlying samples varies considerably, ranging from observations for only 29 students in Meyer (1996) to 1,128 students in Funken *et al.* (1996). While Funken *et al.* (1996) covers girls in single-gender and mixed-gender schools, the other studies compare students in mixed and single-gender classes within coeducational schools. The report by Nyssen (1996) refers to comprehensive schools, while the remaining studies consider students in the highest secondary school track (*Gymnasium*). Evidence is based on outcome measures such as the students' self-assessed interest and motivation. The measured effects are mixed with a tendency to interpret in favour of single-gender education.

Besides the selection bias, these qualitative reports are also likely to suffer from substitution biases: Different treatment measures are applied to different groups (cf. Nyssen, 1996; Volmerg *et al.*, 1996) so that a simple comparison of outcomes is hardly appropriate for identifying an effect of interest. One issue related to this critique is that it is not assured that

[87] Concerning older studies, especially in the 1990s, different publicly financed measures were conducted in order to increase girls' participation in technical subjects and science (including computer science). Such and further measures (supported and funded by the federal commission Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, BLK) primarily focusing on *tertiary* education are summarized in BLK (2002). Additionally, compare Kessels (2002), Hoffmann *et al.* (1997) or Volmerg *et al.* (1996) for brief summaries of older studies for Germany. For the most part, older studies simply compare girls in single-gender and mixed schools and exhibit the same problems as discussed for the studies summarized in Table 4E.1 in the 'Appendix to the Literature Review'. A further stream of literature, which is not the focus of this review, considers gender related subject of degree choice in higher secondary schools (see Roeder and Gruehn, 1997 and Heinrichs and Schulz, 1989 for examples of such studies).

the different groups are taught by similar teachers. Additionally, it is reported in at least one of the four studies (Volmerg *et al.*, 1996) that girls anticipated the expected results so that the measured effects are probably driven by an observation bias or Pygmalion effect (cf. Section 4.2.2).

Concerning non-experimental evaluation studies ('type 2'), the overview article of Mael *et al.* (2005) documents that only few of the existing international, English-language studies control for relevant characteristics driving the selection. Specifically, it seems that the more appropriately selection is controlled for (*i.e.* the more control variables are added) the higher the reduction in the estimated effect of single-gender education.[88] Billger (2006) is an exceptional study using *econometric* techniques in order to take selection to different educational regimes into account. Since it is a recently published paper, it is not included in the 2005 review of Mael *et al.* The paper examines effects of single-gender school attendance on education and individual labor market outcomes in the U.S. Results from regression analyses (including switching regressions for starting salaries) controlling for a variety of students' background characteristics show modest positive effects of single-gender education. However, most regression results are probably biased because of selection in unobserved variables and endogeneity of some regressors. The switching regression results may suffer from the lack of exclusion restrictions in the regime equation, *i.e.* it is not discussed whether variables are included in the regime equation that have no direct potential effect on starting salaries. Consequently, it is not clear whether the study really identifies the causal effect of single-gender education or if the results are still biased due to selection.

The non-experimental evaluation literature ('type 2') for Germany is summarized in the second panel of Table 4E.1. Rost and Pruisken (2000) discuss the selection problem and aim to solve it by comparing similar students. For this purpose they compare samples of fifth and sixth graders (649 students) in single-gender and mixed-gender higher level secondary schools where all three schools considered are similar in that they are run by the Catholic Church. The study finds no significant effect of the organizational class type on different psychological and social outcome variables (including students' self-assessment in mathematics and biology). The contribution of Rost and Pruisken (2000) is that they address the selection problem and possibly even identify a (causal) effect of single-gender education for the relatively limited group of Catholic private school students. However (as is mentioned in this very study), this effect may probably not be generalized for the whole of secondary schools.

[88] The same sceptical conclusion is for example drawn in the German papers by Rost and Pruisken (2000), Faulstich-Wieland (1999) and Baumert (1992) which briefly review the international (and German) evidence on single-gender education.

Furthermore, the authors do not present evidence on the distribution of student background variables among the compared schools. It is not clear whether they really solve the selection problem even for the very specific group of students.

A series of recent publications refers to an intervention study in the German state of *Schleswig-Holstein* in the 1990s (Häußler and Hoffmann, 2002, 1998, 1990 and Hoffmann *et al.*, 1997) and focuses on science (physics) classes. A core sample includes 150 girls and 139 boys in treatment classes from six schools which are taught by six different teachers. Furthermore, 103 girls and 64 boys from two schools are in control classes and are taught by six different teachers. All students considered are seventh graders in the higher level secondary school (*Gymnasium*) in 1992/93. Students are assessed by several standardized written tests and questionnaires during the school year. Differences between groups and difference-in-differences (*i.e.* the development over time between differently treated groups) are calculated. Outcome measures refer to the students' interest, self-concept and achievement. The study may be considered to be a 'type 1' (correlation study) or a 'type 2' study (non-experimental evaluation study) since it is not clear whether the different groups are really comparable in the beginning of the intervention study. There are just brief statements (cf. Häußler and Hoffmann, 2002, page 879 and page 882) suggesting that there were no significant initial differences between groups. Yet, treatment and control groups are located at different schools and may not be comparable. Hoffmann *et al.* (1997) presents regression-adjusted results controlling for students' initial performance and 'learning environment' (as determined by class and teacher, cf. page 149). This regression analysis is likely to suffer from endogeneity of the regressors (*e.g.* self-concept which is used as a control-variable may already be affected by the treatment). The studies conclude that there is a positive impact of single-gender education especially on girls' outcomes.[89]

While correlation studies are frequent and evidence on experimental studies is already rather limited, hardly any truly experimental evidence on single-gender education effects is provided by the literature: As mentioned in Section 4.2.2, few studies draw on natural experiments considering the effects that arise from a transition of schools from a coeducational to a single-gender regime or *vice versa*.[90] Little evidence is documented referring to the transi-

---

[89] Hoffmann *et al.* (1997) additionally provides evidence for chemistry classes. However, due to problems related to the realization of the intervention study (small sample size, deviation of teachers from standardized curricula, cf. page 10), this evidence is difficult to interpret. Previous evidence is additionally presented in Häußler and Hoffmann (1990), observing single-gender physics classes in 1988/89. I do not report on this evidence in detail since the study is rather qualitative.

[90] Marsh *et al.* (1988) conducted such a study for Australia, examining two singe gender schools in the same neighbourhood that were restructured to be coeducational. The study does not find significant effects of the regime change on students' educational performance but positive effects on students' self concept. However, Mael

tion from former single-gender schools to coeducational schools after 1950 in Germany (cf. the third panel of Table 4E.1). Baumert (1992) refers to the regime switch concerning higher secondary schools between 1965 and 1975 and draws on the fact that some schools had already adapted to the new (mixed-gender) regime while other schools had not. Using data from a survey of 12,000 seventh graders including standardized German, English and mathematics outcomes, the study thus compares students in the different school types. Baumert (1992) shows that there is selection into the differently organized school types in regions where both types coexist so that students (or actually their parents) may choose to attend single or mixed-gender schools: Generally, more able students seem to prefer single-gender schools. The further study aims to take this selection into account by analysis of variance controlling for the organizational type of the school. However, the study does not make use of the natural experiment through the potential before-after-comparisons or difference-in-difference estimates. All in all, within the setup of the study the author concludes that both genders perform significantly better in mathematics in single-gender schools than in coeducational schools. Similarly, girls' interest in mathematics decreases notably in seventh grade when girls are taught in mixed-gender classes. No effects are found for the English and German performance.

Concerning social experiments ('type 4'), Kessels (2002) reports on a German intervention project which stands out from the other studies because of an effort to randomly assign students to treatment and control groups.[91] Seven coeducational schools in Berlin participated in this project, where coeducational and single-gender education classes were compared in science (physics and chemistry) lessons. The core sample of the study contains 270 eighth-graders from four comprehensive schools in Berlin (87 girls and 62 boys in coeducational classes as well as 56 girls and 65 boys in single-gender classes). Each participating teacher instructs at least one single-gender and one coeducational class. Information on socio-demographic variables and outcome measures such as motivation and self-concept are assessed by standardized questionnaires and evaluated by analysis of variance. Kessels (2002) finds that there is a positive impact of single-gender education on girls' motivation and self-confidence.

One drawback related to the design of the study by Kessels (2002) is that the study gives no information to verify the assumption that the (randomly constructed) treatment and

et al. (2005) raise some doubts on the study's identification strategy which are mainly related to Pygmalion effects.

[91] Also, from an international perspective there is hardly any evidence based on truly experimental studies. Marsh and Rowe (1996) summarize and re-analyze evidence from studies relating to an experiment in an Australian school. However, the paper raises severe critiques concerning the implementation of the studies and the underlying 'experiment' (*e.g.* related to non-random assignment in the second year of the project).

control groups do not differ (by hazard) in observed characteristics. Related to this, there is significant sample attrition (non-respondence) which might be systematic and thus bias the results. A further problem might be that students are informed about the subject of the intervention study so that the observed *psychological* outcomes are probably biased because of Pygmalion effects. No evidence on hard performance measures is provided. It is not clear whether an increase in girls' perceived motivation related to the science class is accompanied by a real increase in their science knowledge.

To sum up, even if several studies discuss the implication of single-gender education and refer to intervention projects that were conducted in order to shed light on this topic, most work is rather qualitative and forgoes using appropriate statistical identification strategies.[92] In other words, most of the existing studies imply conclusions (mostly in favor of single-gender education) which are rather equivocal from a methodological point of view. The major problem inherent to previous studies is that they do not solve the selection problem arising from the fact that compared single-gender and coeducational groups probably consist of students with different characteristics.

Similarly, most studies suffer from the impossibility of identifying effects from multiple treatments: If these studies try to identify the effect of single-gender education by comparing students in treatment (single-gender) and control (coeducational) groups, where besides the organizational treatment both groups differ by further treatment measures (*e.g.* different teachers, different curricula and educational methods), the effect of the single-gender treatment can hardly be identified. In these cases it is strictly not possible to disentangle the effects of different measures and the presented conclusions of these studies remain speculative in nature. Additionally, most studies refer to psychological variables or social as opposed to hard performance measures. This is probably partly due to the fact that German data protection laws are rather strict, implying that school representatives are not willing to report on their students' grades.

[92] The same conclusion is drawn by Ludwig (2003) who also reviews the literature with a focus on German studies.

4.3 Detailed Facts and Findings from the BW-project

This section provides detailed background information on the BW-school project and its evaluation: The general framework and implementation issues of the project are described in sub-section 4.3.1. Sub-section 4.3.2 discusses the identification strategy and provides evidence on selection of students to different groups. General gender differences emerging among the students considered are addressed in sub-section 4.3.3, while sub-section 4.3.4 presents the evaluation results of different groups. Specifically, overall outcome differences of students (and especially girls taught by the same teacher) in single and mixed-gender groups are shown.

4.3.1 Facts and Implementation of the project

The BW-project was conducted during the school years 2004/2005 and 2005/2006 in a lower level secondary school *(Haupt- und Werkrealschule)* in the city of Rastatt in the state of Baden-Württemberg.[93] The specific school is of intermediate size, consisting of about 370 students in grades five to ten. The intervention refers to computer science lessons of approximately eighty students who are in fifth grade (aged 11-13) in the first year of observation. These computer lessons take part once a week for one hour. Primary goals are to familiarize students with the computer in general (*i.e.* the different components of the hardware) and to teach the application of specific standard software, especially related to text processing and the Internet. While the fifth graders are taught in three separate classes in all subjects, there are six computer study groups. Random assignment of pupils to the different groups was not feasible. The school's principal opposed random assignment explaining that necessities related to the students schedule and the available computer classrooms ruled the design of the computer groups.

In the first year of the project, four of the groups were mixed groups (coeducational) and two groups were single-gendered (one all-boy group and one all-girl group). Table 4.2

[93] Rastatt is located in the higher plain of the river Rhine (Rhein) between the Rhine and the Black Forrest in the direct vicinity of the French region Alsace and the German city of Karlsruhe. The population size amounted to about 47.000 individuals in 2004 (the year when the project started). The population density corresponded to the intermediate population density in the state of Baden-Württemberg. Compared to the entire German population, the population is representative concerning the gender and age structure. In 2004, the unemployment rate in Rastatt amounted to 6.3 % which was below the West German rate of 9.4 % and slightly lower than the state Baden-Württemberg's unemployment rate (6.9 %). Generally, the proportion of persons with an immigrant background (holding citizenships from foreign states) is higher in the state of Baden-Württemberg (12.1 %) than in most other German states (with an average of 8.9 %). Compare Rastatt (2005) and IW (2005) for these and more statistical details.

illustrates the (gender) composition of the six groups in which fifth graders are taught in computer sciences and indicates which teacher is responsible for which group.

Table 4.2: Division of groups in grade 5

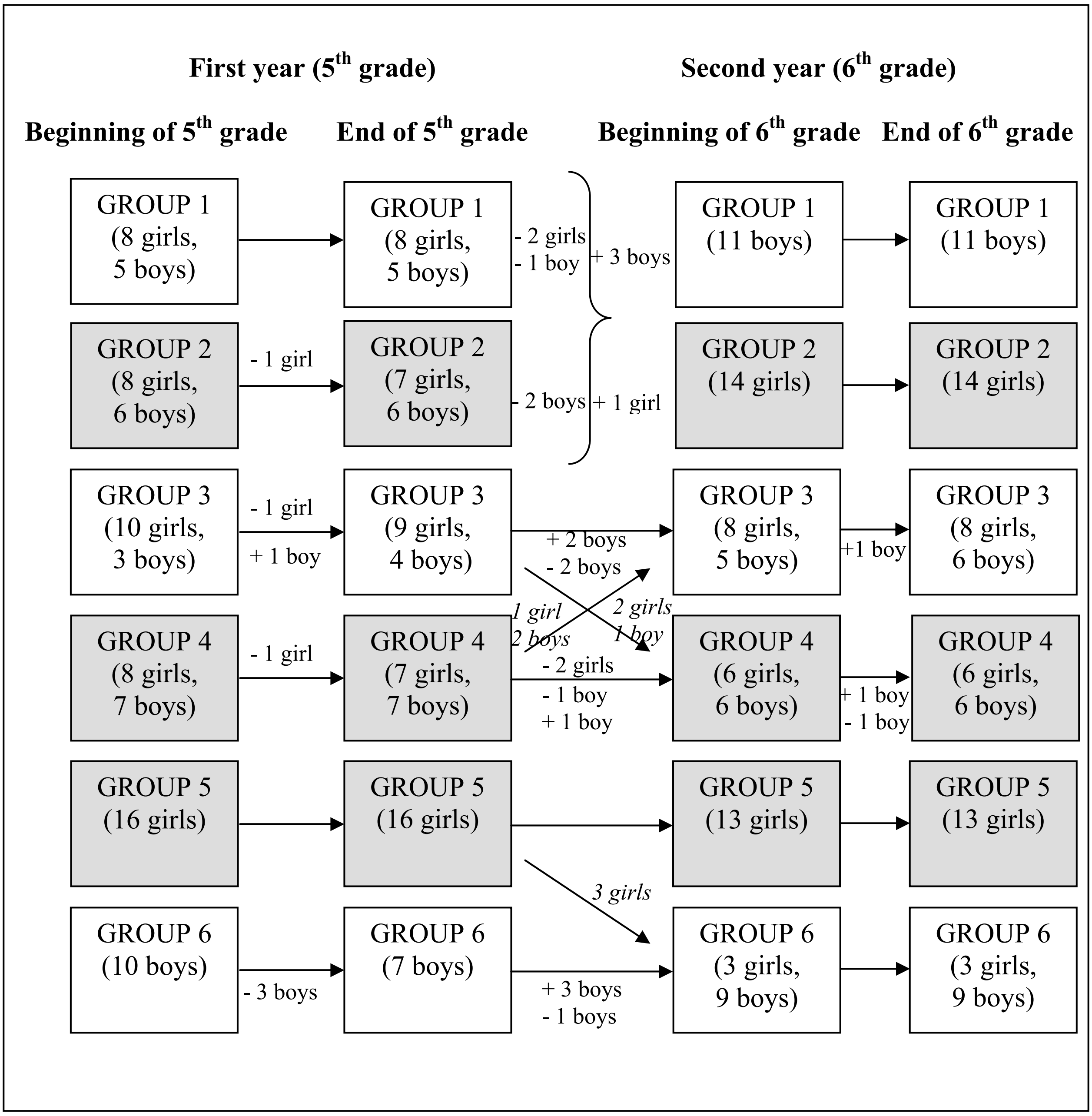
Group	Class	Teacher	Group size*	Number of girls *	Share of females
1	5a	A	13	8	62 %
2	5a	B	13 (+ 1)	6 (+ 1)	50 %
3	5b	A	13 (+-1)	9 (+ 1)	69 %
4	5b	B	14 (+ 1)	7 (+ 1)	50 %
5	5c	B	16	16	100 %
6	5c	C	7 (+ 3)	0	0 %

Note: * (+) indicates the number of students who were in this group at the beginning of the school year but dropped out of class until the end of term. This number of students is not included in the total group size. (-) indicates the number of students who were not observed in the beginning of the school year but joined class during the school year. This number of students is included in the total group size. Grey-shaded lines refer to groups taught by the same teacher (B).

Groups five and six are single-gender groups while the remaining groups are coeducational. The all-girl group (group 5) is taught by the same computer science teacher who instructs groups two and four. Group sizes changed over time due to students moving to other school districts or schools and due to students entering the school from other schools, respectively: Seven students dropped out during the first year while one additional student joined group 3 during the school year.

Figure 4.1 gives an overview of the group compositions by gender at different points of measurement, including the second year of observation. In addition to a switch of groups 1 and 2 from coeducational to single-gender groups, three girls from the former all-girl group 5 were now coeducated in group 6 (the former all-boys group). The latter measure was allowed by the school principal in order to equalize the group sizes and additionally complicates the comparability of groups. Additionally, the teachers allowed some students to switch between groups three and four regardless of the intervention project. Furthermore, all in all, 18 students drop out by the end of the project. Ten students, nine boys and one girl, join the groups as new students in the second year (sixth grade). Taken together, during the first and in the second year of observation three boys (one in the first year and two in the second year) joined the groups. Considering all four measurement times (mid-term and end of term of both years), only 64 students were observed throughout the two-year period.

Figure 4.1: Group compositions by gender



Note: + indicates the number of students who were in this group in the previous term but dropped out until the current term. - indicates the number of students who were not observed in the previous term but joined the group until the current term. Grey-shaded groups refer to students taught by the same teacher (B).

If the sample is restricted to these 64 students, the proportion of females amounts to about two thirds of the students in the resulting sample (43 out of 64). Eight girls are not observed throughout the two-year span. Thus, fluctuations are especially high among boys, *i.e.* 22 boys are not observed throughout the entire time span while 21 boys remain in the same school from the beginning of grade five until the end of grade six. Table 4.3 shows the numbers of continuously observed students and the numbers of students not observed throughout the entire time span by gender and group number.

Table 4.3 : Students observed over time and dropouts

	Continuing Students		Changers	
Group	Male	Female	Male	Female
1	7	0	6	2
2	0	13	2	2
3	4	8	4	1
4	4	6	3	3
5	0	13	0	0
6	6	3	7	0
Total	21	43	22	8

Note: The group number refers to the group indicated in the second year. Continuing students are students observed throughout the two-year time span of the project. Grey-shaded lines refer to students taught by the same teacher (B).

From the above considerations it becomes clear that the time dimension of the project may hardly be used. Due to the unintended switch between groups of some students in the second year and the dropping out of other students, the group of students that might potentially be compared over time becomes too small. Specifically, only few girls are taught by the same teacher in both years: There are seven girls in group 2, six girls in group 4 and thirteen girls in group 5. In other words, too many confounding factors (including teacher differences and group compositions) in combination with the limited and decreasing sample size rule out comparisons over time. Thus, the following analyses are confined to examining the results for the first year of the project.

Student, class and teacher background characteristics are assessed through mainly self-contained questionnaires. Table 4.4 summarizes the different dates of measurement. At the beginning of the project (in January 2005), parents and student characteristics are assessed. Variables collected in these questionnaires are primarily related to individual age, overall school performance and immigrant background, the age and number of siblings, parental education and employment. The assessment of student characteristics is repeated once every six months until the end of the project. Additionally, the teacher questionnaires were distributed at the end of the first project year. Here, further information concerning the different groups as well as teacher characteristics and their gender views are assessed. All the questionnaires are presented in the 'Appendix of Questionnaires' (Appendix 4A). Stylized facts on the general student's characteristics which are drawn from the questionnaires are summarized in the Appendix 4C ('Appendix on Stylized Facts on the Students' Background').

Table 4.4: Dates of measurement and questionnaires

Wave	Date	Questionnaires
1	January 2005 (mid-term)	Parent and student questionnaires
2	June 2005 (end of term)	Student and teacher questionnaires
3	January 2007 (mid-term)	Student questionnaire
4	June 2007 (end of term)	Student questionnaire

Note: Questionnaires for the first year are given in Appendix 4A.

Besides the control variables, outcome variables of interest need to be collected. As indicated in Section 4.2, there are different kinds of performance measures which may be categorized as 'hard' and 'soft' measures. Hard performance measures directly relate to the educational performance and may be assessed by grades, students' self-assessment of their own performance or by test scores. Direct assessment of performance via students' grades might be problematic because grades are likely to be relative instead of absolute performance measures, *i.e.* they indicate the students' performance *within* a given group. Students' self-assessment might be problematic as well (especially in the gender context) because girls tend to understate their own performance (compare Section 4.1).

Thus, it is more reliable to base performance statements on standardized students' tests.[94] In the BW-project computer tests are conducted at two points in time in order to assess students' performance over time: about two weeks before mid-term and two weeks before the distribution of end of term school reports. The test questions are shown in the 'Appendix of Tests' (Appendix 4B). These tests are not a part of the students' term grade, yet the students are not aware of this which is a promising strategy in order to assure that the students put the required efforts and seriously answer the questions. The contents of the examinations were jointly developed by all of the three participating computer science teachers with the objective of not giving an advantage to one of the tested groups. All teachers graded the tests according to a linear scale. Even if the assessments are rather short, according to the teachers they yield an overall measure which is generally representative for the students' real performance.

In addition to standardized test outcomes, the BW-project draws on 'comparative' hard performance measures which relate to the students rating of their own performance relative to the other students. The underlying scale ranges from 1 ("I perform much better than my classmates.") to 5 ("Other students perform much better than me."). A value of 3 indicates that the student thinks that she performs equally to the average student. This performance measure is provided for the overall performance in all subjects taken together, for computer science, for math and for the German class. Note that the underlying scale is not an absolute performance measure but a relative self-ranking of students towards their classmates. At the

[94] For example Hoffmann *et al.* (1997) contains detailed recommendations for the construction of such tests.

same time it is likely that the indicated value is influenced by the grade the student achieves in the respective subject.

Furthermore, students are asked about the grades they expect to achieve in the end of term school report. Generally, the German school grades follow a scale from 1 ("outstanding") to 6 ("failed"). According to the school principal's information, in this school all grades are usually based on a linear scale where the distances between two grades are proportional to the performance difference (for examples related to the test scores in underlying examinations).[95]

As mentioned above, besides these different 'hard performance' measures, soft performance measures, as they are often used in sociological studies, are observed as well.[96] These measures relate to the students' gender perception and motivation. Motivation is assessed by the question whether the student likes working with the computer. Gender perceptions are deduced from the answers to the question whether the student thinks that boys or girls (or both) know more about computers relative to the other gender. In addition, students are directly asked whether they prefer to be taught in mixed or single-gender groups.

4.3.2 Identification Strategy and Selection Issues

Given the non-experimental set-up of the school project (*i.e.* no random assignment to treatment and control groups), the evaluation calls for statistical methods in order to take possible selection effects into account. However, popular evaluation techniques like matching or regression analysis crucially depend on the given sample size: Degrees of freedom drop if more control variables are included (*i.e.* the more appropriately selection is corrected) and it might be hard to identify significant effects if few students are observed in the treatment and control groups. Specifically, in the present case sample sizes are too small in order to employ such common methods. Too few comparable girls (*i.e.* those taught by the same teacher) are observed in mixed and single-gender classes respectively. Additionally, it is hard to impose parametrical assumptions on the data, which would be required for simple standard statistical tests (*e.g.* the t-test).

Therefore, the analysis opts for a feasible solution and proceeds as follows: First of all, it is demonstrated that students in different groups do not - in fact - differ in their observed characteristics in the given case. Based on this insight, students in single and mixed-gender

[95] Thus, in this regime, grades can be considered to be metrically scaled, which is important when discussing the results from the project.
[96] For the interested reader, the development of detailed soft performance scales is extensively discussed in Kessels (2002, chapter 5.5) or Hoffmann *et al.* (1997, chapter 6.2), for example.

groups taught by the same teacher are compared using non-parametrical χ^2-homogeneity-tests or Kruskal-Wallis tests. The advantage of these statistics is that they are not observation-intensive but may be applied to the small samples.

Specifically, for the nominal and ordinal variables standard χ^2-independency tests are used to assess whether the characteristics are similarly distributed within groups. The conducted independency test may be interpreted as a homogeneity test with the null hypothesis that the different groups are drawn from the same population.

For metrical variables Kruskal-Wallis rank tests are feasible in order to examine whether the samples (*e.g.* the six different groups) are drawn from the same population:[97] The null hypothesis assumes that the population medians are equal (H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_6$, H_1: $\mu_i \neq \mu_j$ for at least one set of i and j, where $i \neq j$ *and* $i,j \in [1,2,3,\ldots,6]$.). The test statistic is based on ranking the combined sample of all observations. Then, the sum of the ranks is computed for each of the groups (R_1 to R_6). The test statistic is:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)\,, \qquad (12)$$

where n_i $(i = 1, 2, \ldots, k)$ represents the sample size for each of the k groups. The intuition behind this test statistic is that if the groups really have the same median, the sum of ranks for each group should not differ too much. If the null hypothesis of equal populations is true, this statistic is approximately χ^2-square distributed with k-1 degrees of freedom. The distributional assumption is valid if each of the n_i is at least five.

Distribution of Group Background Variables

Detailed contingency tables of group characteristics and means of the background variables by group are presented in Appendix 4D ('Appendix of Tables of Group Characteristics'). Furthermore, Table 4.5 presents the results from χ^2-independency tests for the nominal and ordinal variables. Since the tests require at least five observations per cell (otherwise the distribution of the test statistic cannot be approximated by the χ^2-distribution), groups are aggregated into coeducational and mixed-gender groups. Table 4.5 shows that the null hypothesis (for each of the observed variables) cannot be rejected.

Table 4.6 reports the Kruskal-Wallis test results for observed metrical variables. According to these results, one cannot reject (at any conventional level of significance) the assumption that the samples are drawn from the same population.

[97] The test is developed in Kruskal and Wallis (1952).

Table 4.5: χ2-tests of group homogeneity (dummy variables)

Variable	$\chi^2(1)$	Probability
German-born	0.03	0.87
German-born parents	0.21	0.65
German language spoken at home	0.23	0.63
Some immigrant background	0.35	0.55
Grandparents live nearby	0.05	0.83
Higher secondary education of parents	0.01	0.92
Parental interest in school affairs	0.01	0.90

Note: The χ^2 –test compares distributions of coeducational and single-gender groups. A comparison on a less aggregated level (*i.e.* for the six study groups) is not possible due to the limited sample size.
Source: Student questionnaires. Own calculations.

Table 4.6: Kruskal-Wallis tests by group for observed metrical variables

Variables	$\chi^2(5)$	Probability
Age at first measurement date	0.75	0.98
Number of siblings at home	4.17	0.52
… siblings more than 3 years younger	1.28	0.94
… siblings (less than) 3 years younger	2.25	0.81
… siblings (less than) 3 years older	1.84	0.87
… siblings more than 3 years older	0.49	0.99

Note: The test statistic is explained in Section 4.3.2.
Source: Student questionnaires. Own calculations.

Thus, all the presented findings provide confirmation that the groups do not differ considerably in their observed characteristics. However, for the following analyses one needs to bear in mind that only three of the groups considered are taught by the same teacher. Comparisons among these groups are expected to be especially meaningful.

Teachers' statements

In addition to the above considerations, information from the teacher questionnaire is used in order to determine (1) whether the groups are comparable according to the teachers' opinion and (2) whether there are teacher differences concerning gender views. The teacher questionnaire is given in the 'Appendix of Questionnaires' (Appendix 4A).

Table 4.7 shows that the teachers' estimations on the proportion of students with an immigrant background in each group differ from the true proportions (as indicated by the individual students). This probably stems from the fact that teachers are only vaguely informed about the students' family background. According to information from the school principal the parents' participation in school events like parent-teacher conferences is rather low (or practically non-existent) which might make it more difficult for teachers to judge students' family background. The share of immigrant children is underestimated in groups 1, 3 and 6 and overestimated in the other half of the groups (all of which are instructed by teacher B).

Generally, (as discussed in Section 4.2.4.1) the share of immigrant children is high in each group (at least 50 %). Table 4.7 includes children speaking a foreign language at home in the definition of immigrant children. Based on this definition drawn from the student-questionnaires, the shares of immigrants are between 69% and 92% depending on the group.

According to the teachers, in each group there are at most two students with serious language problems and up to one student is classified as showing behavioural or learning problems. There seems to be no clustering of disadvantaged or problem-children into one group or to one teacher. In addition, the questions related to the teachers' educational treatment of the class show that there is some homogeneity: All teachers but one (teacher C of group 6) assign homework less than once a month. Teacher A (groups 2 and 4) also instructs the religion and social study classes of the children in his computer group. It is to be expected that the homogeneity of the educational treatment is the largest between the groups taught by the same teacher B, where this teacher does not meet one of his groups in any other class.

Table 4.7: Teachers' information on group background

Variable / Group	1	2	3	4	5	6
Share of foreigners (born abroad or parents born abroad)	38 %	100 %	67 %	86 %	100 %	57 %
Real share of foreigners (student information)	*69 %*	*75 %*	*92 %*	*50 %*	*69 %*	*86 %*
Share (number) of children showing serious language problems	15 % (2)	0 %	0 %	14 % (2)	0 %	0 %
Share (number)of children showing learning / behavioural problems	8 % (1)	0 %	0 %	7 % (1)	6 % (1)	4 % (1)
Frequency of homework in a month*	<1	<1	<1	<1	<1	1-2
Number of other subjects in which teacher instructs these students	2	0	1	0	0	0
Kind of other subjects in which teacher instructs these students	social studies		social studies			
Teacher ID	A	B	A	B	B	C

Note: <1 'sometimes but less than once a month', 1-2 'once or twice a month, but less than once in two weeks'.
Source: Teacher questionnaires. Own calculations.

Background information on the teachers is given in Table 4.8. All the teachers are male (which might be important concerning their gender views). Teachers A and B are more similar in their age, work experience and instructed subjects: They are relatively young (36 and 37 years respectively), teach for three and eight years respectively and generally instruct children in all kinds of subjects, while teacher C is a 56 year old science teacher with 33 years of work experience. From this and the previous tables it becomes obvious that group six (instructed by teacher C) is hardly comparable to the other groups if one expects that differences in teacher characteristics influence the educational treatment.

Table 4.8: Teacher characteristics

Variable / Teacher	A	B	C
Gender	male	male	male
Age	36	37	56
How many years have you been a teacher?	3	8	33
Number of subjects taught	6	9	4
Kind of subjects taught	*all fields**	*all fields**	science

Note: **all fields* indicates that the teacher instructs science/math as well as languages and social studies.
Source: Teacher questionnaires. Own calculations.

Teachers' attitudes towards gender views and stereotypes have been inquired into through a series of questions which are repeated in Table 4.9. Teachers B and C, who have gained experience with single-gender classes in the first project year, are consistent with the view that it is reasonable to segregate pupils into single-gender classes. According to teacher B who teaches two coeducational and one all-girl computer class, girls benefit from single-gender education because they interact more in class compared to the situation in a mixed-gender classroom where they ask fewer questions. Teacher C who teaches the all-boy class thinks that boys generally believe to be more talented in science than girls and therefore debar girls from actively participating in science classes. Teacher A states that he cannot judge the issue since he does not have any experience with single-gender classes. The indetermination of teacher A might also be attributable to his relatively few years of work experience.

Generally, all three teachers do not think that one gender is more talented in computer studies compared to the other gender. They are uniformly of the opinion that girls are more talented in learning languages than boys, while boys are not more talented in technical subjects than girls (teacher B is unsure concerning the second point). None of the teachers expresses objections to girls being talented enough to work in technical professions and only teacher A thinks that mainly boys rather than girls should seek technical professions. The last statement might be considered to be consistent with the statement that girls are more talented in learning languages.

The teachers think that different strengths and weaknesses of girls and boys in different school subjects are *both* instilled by society and innate, while the first source is considered to be of greater importance. Comparing boys' and girls' overall performance in lower secondary school, teachers A and C state that boys have more problems keeping up with the educational contents than girls. Teacher B thinks that none of the genders has more or less problems compared to the other.[98]

[98] In addition to the facts presented, the questionnaires provide the information that all teachers agree that the school is well equipped with computers and software.

Table 4.9: Teachers' gender views and stereotypes

Question / Response of teacher	A	B	C
What do you think …			
… Is it wise to teach boys and girls separately in the computer class?	do not know	yes	yes
… Are boys or girls more talented with respect to computer studies?	neutral	neutral	neutral
According to your professional experience, which of the following statements are correct?			
… Girls are more talented than boys when it comes to learning languages.	yes	yes	yes
… Boys are more talented than girls when it comes to learning maths or science.	no	yes/unsure	no
… Girls have no talent for technical professions.	no	no	no
… Boys rather than girls should seek technical professions.	yes	no	no
… Different strengths and weaknesses are instilled rather than innate.	yes	yes	yes
… Different strengths and weaknesses are both instilled and innate.	yes	yes	yes
… On average, girls have more problems keeping up with the secondary school level's curriculum than boys.	no	no	no
… On average, boys have more problems keeping up with the secondary school level's curriculum than girls.	yes	no	yes

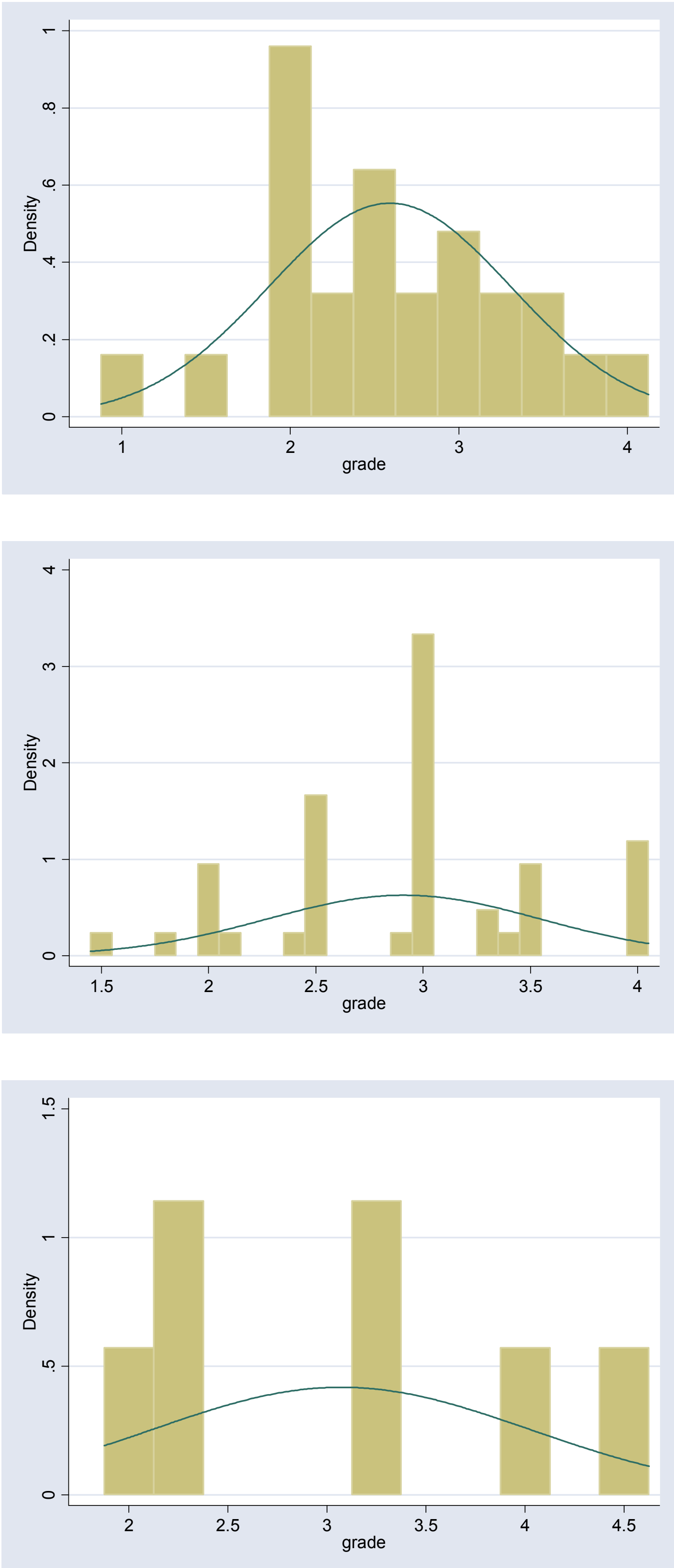
Source: Teacher questionnaires. Own calculations.

All in all, the answers do not indicate that one of the teachers is affected by serious 'traditional' gender stereotypes. However, one has to bear in mind that the teachers were informed about the contents of the school project. Thus, it cannot be ruled out that their answers are biased in the direction of the answers they expected to be politically correct.

Performance by teacher

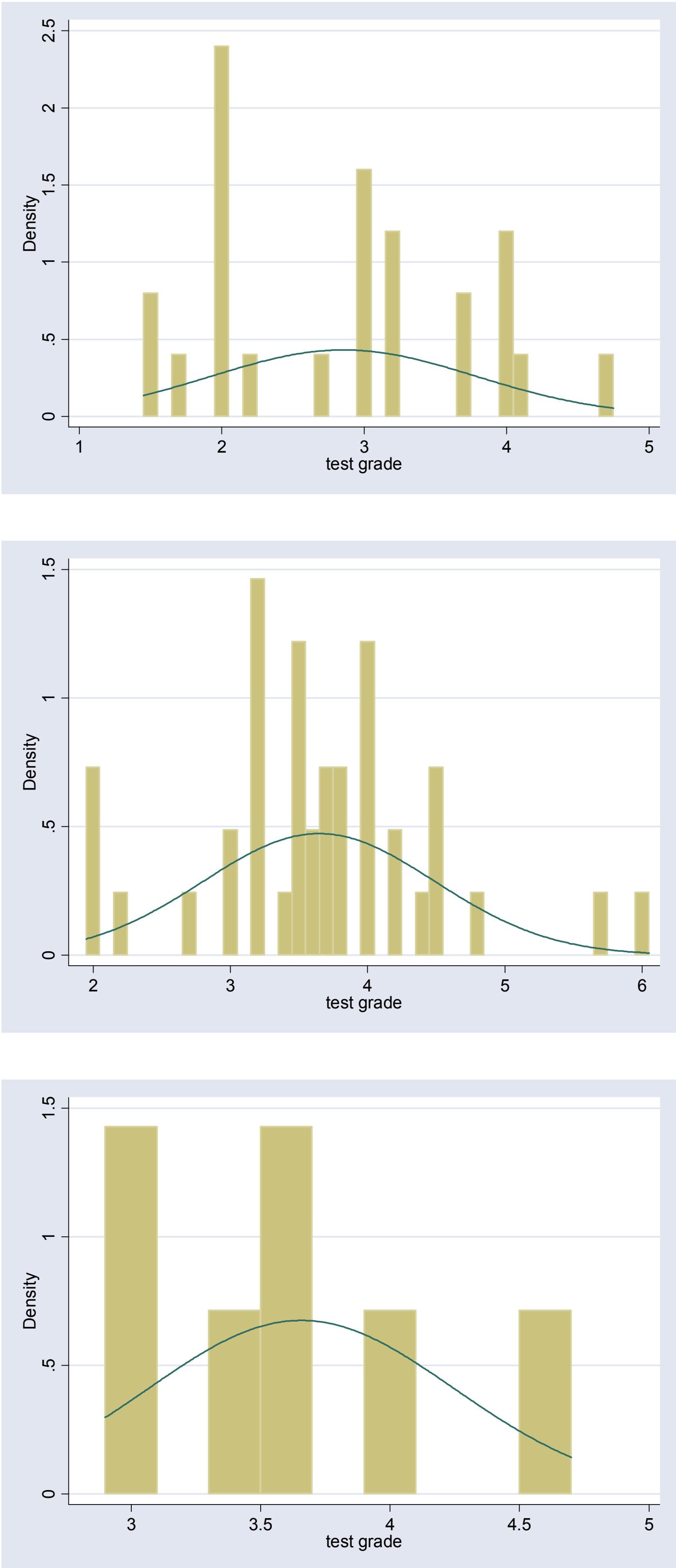
Even if the different groups are similarly composed, the students' outcomes will hardly be comparable between groups if teacher quality varies between groups. Figure 4.2 shows the distributions of grades by teacher in the computer test at mid-term (together with a sample normal distribution). Figure 4.3 refers to the computer tests at the end of term. Further statistics related to the distribution of grades by teacher are given in Table 4.10.

Figure 4.2: Distribution of mid-term computer test grades by teacher



Note: The distributions refer to teachers 1 to 3 respectively (from top to bottom).
Source: Test results. Own calculations.

Figure 4.3: Distribution of end of term computer test grades by teacher



Note: The distributions refer to teachers 1 to 3 respectively (from top to bottom).
Source: Test results. Own calculations.

It is obvious that the grade distributions differ substantially by teacher. While, for example the mode of the distribution related to teacher A at mid-term is grade 2, it is grade 3 for teacher B. The corresponding Kruskal-Wallis test related to the end of term grade indicates that one can reject the hypothesis of equality of populations at any conventional level of significance.[99] Since the teachers use the same scale in order to score the tests[100] and given that the socio-economic background of the students is very similar among groups, it is likely that these differences are not only driven by differences due to the single-gender and coeducational treatment but also by teaching differences.

Table 4.10: Distribution of grades by teacher

Teacher	A	B	C
Test grades at mid-term			
Mean	2.59	2.91	3.07
(s.d.)	0.72	0.64	0.95
Median	2.50	3.00	3.25
Mode	2.00	3.00	2.25, 3.25
Skewness	0.00	-0.07	0.29
Kurtosis	2.59	2.52	1.67
Kruskal-Wallis $\chi^2_{(2)}$		3.42	
Test grades at end of term			
Mean grade	2.86	3.65	3.66
(s.d.)	0.93	0.84	0.59
Median grade	3.00	3.60	3.70
Mode grade	2.00	3.20	3, 3.7
Skewness	0.18	0.33	0.53
Kurtosis	1.89	4.06	2.50
Kruskal-Wallis $\chi^2_{(2)}$		10.48	
Number of observations	25	41	7

Note: The Kruskal-Wallis test refers to the comparisons of the respective distributions for students instructed by different teachers.
Source: Test results. Own calculations.

To sum up, the descriptive evidence presented in this sub-section indicates that the groups are similar in their socio-economic background while they are principally educated in a similar way by the different teachers. However, there seem to be notable teacher differences in grading. Therefore, the following analysis needs to focus on comparisons of students taught by the same teacher to guarantee that similar students (under similar conditions) are considered.

[99] The Kruskal-Wallis test related to the mid-term grade can only be rejected at the 18%-level of significance.
[100] There are minor differences in the valuation of the first test, where teacher 2 uses a more exact scale based on intervals of first digits, *i.e.* [1.0; 1.1; 1.2; … ; 5.9; 6.0] while teacher 1 and 2 refer to intervals of quarter s of grades, *i.e.* [1.00; 1.25; 1.50; … ; 5.75; 6.00]. For the end of term test all teachers use the first digit scale.

4.3.3 Gender Related Findings

Before addressing the issue of coeducation, this section sheds light on the question whether there are overall performance differences by gender among the students observed in the intervention study. First of all hard performance measures related to different subjects are considered. Later on, soft performance measures are compared by gender.

Table 4.11 presents mean grade and performance measures separately by gender together with Kruskal-Wallis tests on the equality of the gender specific distributions. Kruskal-Wallis tests are appropriate if it might be assumed that the scales are metrical. Again, this assumption is reasonable for the grades in this case because of an underlying linear scale. The remaining scales can be taken to be metrical as well if one assumes that students' judge their own relative performance based on their grades.

Table 4.11: Performance in general subjects by gender

Group		All students		Girls		Boys	
Variable	Kruskal-Wallis $\chi^2_{(1)}$	Obs	Mean (s.d.)	Obs	Mean (s.d.)	Obs	Mean (s.d.)
Performance measured at mid-term:							
Overall comparative performance	0.34	74	2.70 (0.77)	46	2.76 (0.57)	28	2.61 (1.03)
Comparative performance in math	3.96**	74	2.80 (0.89)	46	2.96 (0.79)	28	2.54 (1.00)
Comparative performance in German	0.16	74	2.95 (0.90)	46	2.89 (0.88)	28	3.04 (0.96)
Performance measured at end of term:							
Overall comparative performance	0.46	73	2.84 (0.60)	45	2.89 (0.61)	28	2.75 (0.59)
Expected average grade (all subjects)	0.59	67	2.77 (0.79)	40	2.69 (0.81)	27	2.89 (0.76)
Comparative performance in math	3.83*	74	2.84 (0.79)	46	2.98 (0.68)	28	2.63 (0.91)
Expected math grade	2.47	73	3.05 (0.90)	46	3.19 (0.87)	27	2.82 (0.92)
Comparative performance in German	0.16	74	3.07 (0.75)	46	3.07 (0.68)	28	3.09 (0.86)
Expected German grade	1.41	74	3.28 (0.78)	46	3.19 (0.72)	28	3.44 (0.86)
Performance change[+]							
Change in overall comparative performance	0.00	73	0.12 (0.71)	45	0.11 (0.61)	28	0.14 (0.85)
Change in comparative math performance	0.05	74	0.05 (1.06)	46	0.02 (0.95)	28	0.09 (1.23)
Change in comparative German performance	0.79	74	0.13 (0.92)	46	0.17 (0.68)	28	0.05 (1.23)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for boys and girls. ** Significant at the five percent level. * Significant at the 10 percent level.
[+] The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

According to the comparative performance measures in Table 4.11, the average girl attributes a worse relative position to herself as compared to her classmates than the average boy does. This is especially true for math. However, for the German performance the reverse is true. These findings are true for both measurement dates, while the average person of the group indicating a better self-ranking (*i.e.* boys for their general performance and math, and girls for German) experiences a larger drop in his or her self-ranked position. Concerning expected end of term grades, the average girl generally (and especially in German) performs better than the average boy, while in math the reverse is true. The latter finding may explain the higher self-ranking of the average boy concerning his math performance. According to the Kruskal-Wallis tests, the hypothesis of equality of populations cannot be rejected for all available measures, but the comparative math performance measure can. Thus, one might assume that boys in the observed sample generally rate their relative math performance better than girls.

Table 4.12: Computer science performance by gender

Group		All students		Girls		Boys	
Variable	Kruskal-Wallis $\chi^2_{(1)}$	Obs	Mean (s.d.)	Obs	Mean (s.d.)	Obs	Mean (s.d.)
Performance measured at mid-term:							
Computer test grade	1.71	74	2.82 (0.71)	46	2.72 (0.68)	28	2.97 (0.73)
Comparative performance in computer science	1.27	73	2.41 (0.85)	45	2.49 (0.76)	28	2.29 (0.98)
Performance measured at end of term:							
Computer test grade	2.95	73	3.38 (0.92)	45	3.22 (0.92)	28	3.63 (0.89)
Comparative performance in computer science	0.52	74	2.74 (0.66)	46	2.80 (0.58)	28	2.64 (0.78)
Expected computer science grade	0.00	70	2.26 (0.67)	44	2.21 (0.52)	26	2.35 (0.87)
Performance change:[+]							
Change in computer test grade	0.62	73	0.54 (0.96)	45	0.47 (0.93)	28	0.65 (1.00)
Change in comparative performance	0.23	73	0.34 (0.77)	45	0.33 (0.64)	28	0.36 (0.95)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for boys and girls. ** Significant at the five percent level. * Significant at the 10 percent level.
[+] The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Table 4.12 additionally shows performance measures specifically related to the computer class. Concerning the mid-term and end of term computer tests, the average girl performs slightly better than the average boy. The same is true for the expected end of term grade. Even though, the average boy ranks himself on a relatively higher position compared to his class-mate than the average girl does. This may be a hint for a higher self-esteem of boys. However, all the differences are not substantial. The Kruskal-Wallis tests indicate that it is not possible to reject the hypothesis of equal populations for any of the available variables.

Table 4.13 shows the results related to the soft performance measures together with χ^2-homogeneity tests.[101] Most students state that they like working with computers when they are asked around mid-term. Only 9 % of the responding female students (4 out of 42) and 18 % of the male students (5 out of 23) do not like computer work. Based on the χ^2-test, it is not possible to reject the equality of the distribution of outcomes for boys and girls. At the end of term the proportion of students disliking computer work is larger and especially high among male students (46 % *vs.* 20 % of responding female students). The χ^2-test now rejects equality of the male and female distributions on the one-percent level of significance.

Table 4.13: Soft-performance measures by gender

Group		All students		Girls		Boys	
Variable	$p[\chi^2_{(1)}]$	Yes	No	Yes	No	Yes	No
Variables measures at mid-term:							
I like working with computers.	0.24	65	9	42	4	23	5
Boys know more about computers than girls.	0.00	16	56	0	45	16	11
Girls know more about computers than boys.	0.08	9	63	8	37	1	26
Both genders know about computers.	0.00	47	25	37	8	10	17
I prefer to be in a single-gender computer group.	0.29	32	41	18	28	14	13
Variables measures at end of term:							
I like working with computers.	0.01	52	22	37	9	15	13
Boys know more about computers than girls.	0.00	11	61	1	44	10	17
Girls know more about computers than boys.	0.05	6	66	6	39	0	27
Both genders know about computers.	0.04	55	17	38	7	17	10
I prefer to be in a single-gender computer group.	0.00	17	55	17	29	0	26

Note: The χ^2-tests tests refer to the comparisons of the respective distributions for boys and girls. $p[\chi^2_{(1)}]$ indicates the level of significance.
Source: Student questionnaires. Own calculations.

The majority of students think that both genders are equal or similar as concerns their computer knowledge. However, girls' and boys' statements differ with respect to this question: At mid-term none of the girls think that boys are superior concerning their computer skills while the majority of male students (59 %) state that boys know more about computers. These notions converge somewhat around the end of term when one girl (2 %) indicates that boys know more about computers but only 37 % of the boys take this position. The proportion of students thinking that girls know more about computers is rather low: 17 % (13 %) of girls take this position at mid-term (end of term) and only one boy agrees at mid-term and none of the boys at the end of the term. While at mid-term boys most often state that they know more than what girls know about computers, most girls (82 %) assume that both genders know

[101] For the purpose of the χ^2-statistics, some outcome measures are recoded to be binary in order to assure that there are enough observations per cell so that the approximation of the test statistic is valid.

about the same as girls. The proportion of boys sharing the latter opinion rises at the end of term (from 37 % to 63 %).

Around mid-term 44 % of all students, 39 % of the girls and even 52 % of the boys, would prefer to be taught in a single gender class. At the end of term, the proportion remains similar for girls (37 %) while, surprisingly none of the boys wants to be segregated. It is hard to interpret the reasons for the change in the boys' opinion concerning the institution of an all-boys class. Potential reasons might be related to the specific treatment of the teacher who instructed the all-boys class.

4.3.4 Main Results: Group Related Findings

This section examines whether there are feasible group differences related to the outcome measures. First of all, results related to hard performance measures are discussed in detail. Beside evaluation based on Kruskal-Wallis tests some evidence from regression analysis is presented. However, due to the impossibility of controlling for a variety of background characteristics because of the limited sample size, both techniques yield (by definition) the same results. Later on in this section, evidence related to the soft performance measures is presented.

Table 4.14: Computer science performance by group type

Group Type		Mixed Gender Groups						Single Gender Groups					
Sample	all	all		girls		boys		all		girls		boys	
Variable	K-W. $\chi^2_{(1)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Test grade	2.31	51	2.72 (0.70)	30	2.57 (0.69)	21	2.94 (0.67)	23	3.02 (0.71)	16	3.00 (0.61)	7	3.07 (0.95)
Comperat. performance	2.52	51	2.27 (0.78)	30	2.37 (0.72)	21	2.14 (0.85)	22	2.73 (0.94)	15	2.73 (0.80)	7	2.71 (1.25)
Performance measured at end of term:													
Test grade	0.79	50	3.32 (0.99)	29	3.11 (0.95)	21	3.61 (0.99)	23	3.50 (0.77)	16	3.43 (0.84)	7	3.66 (0.59)
Comperat. performance	0.24	51	2.75 (0.63)	30	2.80 (0.41)	21	2.67 (0.86)	23	2.74 (0.75)	16	2.81 (0.83)	7	2.57 (0.53)
Expected grade	0.56	47	2.33 (0.71)	28	2.26 (0.57)	19	2.42 (0.89)	23	2.13 (0.57)	16	2.13 (0.43)	7	2.14 (0.85)
Performance change:$^{+}$													
Change in grade	0.08	50	0.57 (0.99)	29	0.50 (0.93)	21	0.67 (1.07)	23	0.47 (0.90)	16	0.43 (0.96)	7	0.59 (0.78)
Change in comp. perf.	4.1**	51	0.47 (0.76)	30	0.43 (0.63)	21	0.52 (0.93)	22	0.05 (0.72)	15	0.13 (0.64)	7	-0.14 (0.90)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in single-gender and mixed groups. ** Significant at the five percent level. * Significant at the 10 percent level. $^{+}$ The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

A simple comparison of outcomes by group types may give a first hint at whether single-gender education is effective. Table 4.14 shows means of hard performance measures separately for students in single-gender and mixed groups. The numbers are provided separately by gender. Overall, there are no substantial performance differences between students in single-gender and mixed groups. On average, test grades of students in mixed-gender groups are better than those of students in single-gender groups and this is especially true for girls: At mid-term (end of term) the average girl in a mixed-gender group performs about 0.4 (0.3) grade points better than the average girl in a single-gender group. Additionally, the average students' comparative performance ranking is better for the average student educated in a mixed-gender group than for an average student from a single-gender group when measured at mid-term. However, the average end of term comparative performance measure takes about the same value for both groups and the average of the expected end of term grade is 0.2 grade points better for the single-gender group.

Table 4.15: Computer science performance by group

Group		Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
Variable	K-W. $\chi^2_{(5)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Computer test grade	4.88	13	2.58	12	2.98	12	2.60	14	2.75	16	3.00	7	3.07
			(0.70)		(0.62)		(0.78)		(0.70)		(0.61)		(0.95)
Comparative performance	6.51	13	2.08	12	2.42	12	2.00	14	2.57	15	2.73	7	2.71
			(0.95)		(0.67)		(0.60)		(0.76)		(0.80)		(1.25)
Performance measured at end of term:													
Computer test grade	17.08**	13	3.28	12	4.11	12	2.40	13	3.49	16	3.43	7	3.66
			(0.80)		(0.97)		(0.85)		(0.56)		(0.84)		(0.59)
Comparative performance	2.28	13	2.62	12	2.92	12	2.67	14	2.79	16	2.81	7	2.57
			(0.65)		(0.29)		(0.78)		(0.70)		(0.83)		(0.53)
Expected computer grade	6.38	12	2.39	11	2.59	12	2.38	12	1.98	16	2.13	7	2.14
			(0.64)		(0.63)		(0.96)		(0.47)		(0.43)		(0.85)
Performance change:$^{+}$													
Change in test grade	11.15**	13	0.70	12	1.13	12	-0.20	13	0.65	16	0.43	7	0.59
			(0.94)		(1.00)		(0.88)		(0.73)		(0.96)		(0.78)
Change in comparative performance	5.86	13	0.54	12	0.50	12	0.67	14	0.21	15	0.13	7	-0.14
			(0.52)		(0.80)		(0.89)		(0.80)		(0.64)		(0.90)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in different groups. ** Significant at the five percent level. * Significant at the 10 percent level. $^{+}$ The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Table 4D.4 in the 'Appendix of Tables of Group Characteristics' additionally shows means of performance measures related to other subjects. There are no substantial performance differences between students in single gender and mixed gender groups in the first term. It is only in the second term that some of the performance measures indicate that students in the mixed groups perform worse than students in the single gender groups. It is not possible to identify whether this difference is an outcome related to the treatment in the mixed gender groups. Table 4.15 provides means of performance variables related to the computer science class for all six groups. Average test grades around mid-term are similar for the different groups as it is confirmed by the Kruskal-Wallis test. However, as a matter of fact and according to the Kruskal-Wallis test the average end-of-term grades differ substantially between groups with the best average result for group 3 (2.40) and the worst for group 2 (4.11). The mean test grade of the all-girl group takes a value of 3.43 which is quite close to the overall average (3.38).

As discussed above, the huge differences in grades are driven by the different teachers instructing different groups. Because of the teacher differences, it is reasonable to specifically compare the all-girl group to the two other groups instructed by the *same* teacher. The results related to the corresponding groups are shaded grey in Table 4.15. In fact, there might be a hint that single gender education is effective: The average end of term performance is much worse in group 2 and slightly worse in group 4 (3.49) compared to the single gender group 5. However, conducting a Kruskal-Wallis test for the relevant groups reveals that one cannot reject the hypothesis of equal populations up to the 19 %-level of significance.

Table 4.16: Girls' computer science performance by group type

Group		Group 1		Group 2		Group 3		Group 4		Group 5	
Variable	K-W. $\chi^2_{(4)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:											
Computer test grade	5.67	8	2.38 (0.57)	6	2.87 (0.73)	9	2.61 (0.86)	7	2.50 (0.58)	16	3.00 (0.61)
Comparative performance	4.40	8	2.38 (0.92)	6	2.67 (0.52)	9	2.00 (0.71)	7	2.57 (0.53)	15	2.73 (0.80)
Performance measured at end of term:											
Computer test grade	10.44**	8	2.90 (0.69)	6	4.07 (1.01)	9	2.57 (0.93)	6	3.25 (0.56)	16	3.43 (0.84)
Comparative performance	0.32	8	2.75 (0.46)	6	2.83 (0.41)	9	2.78 (0.44)	7	2.86 (0.38)	16	2.81 (0.83)
Expected computer grade	2.38	7	2.24 (0.48)	6	2.58 (0.66)	9	2.22 (0.51)	6	2.03 (0.65)	16	2.13 (0.43)
Performance change: +											
Change in test grade	6.28	8	0.53 (0.98)	6	1.20 (0.95)	9	-0.04 (0.89)	6	0.58 (0.49)	16	0.43 (0.96)
Change in comparative performance	4.32	8	0.38 (0.52)	6	0.17 (0.75)	9	0.78 (0.67)	7	0.29 (0.49)	15	0.13 (0.64)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in different groups. ** Significant at the five percent level. * Significant at the 10 percent level. + The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students. Results are not shown for group 6 because there are no girls in this group.
Source: Student questionnaires, test results. Own calculations.

Concerning the comparative performance measure, girls in the single-gender group 5 tend to perform at the lower end of the group averages at mid-term and close to the overall average when measured at the end of term. For the expected end of term computer grade, girls in the single-gender group also perform close to the overall average.

The above analysis refers to both boys and girls. However, in order to measure the effect of education in the all-girls group, it might be more appropriate to compare their performance to the performance of other *girls* in the remaining groups. Therefore, Table 4.16 shows the averages of performance measures of girls only.

The averages and Kruskal-Wallis test presented in Table 4.16 confirm the findings from Table 4.15. Again, there are substantial differences concerning the distribution of the end-of- term computer test grade. However, the effect of single gender education is not clear because at the end of term girls educated in the single gender group (group 5) perform only better than girls educated in *one* of the two coeducational groups educated by the same teacher (group 2) but worse than girls in the other group (group 4) if test grades and the expected computer grade are considered.

Table 4.17 shows averages of the performance measures for boys in each group. However, the group sizes drop to very small numbers so that it is hard to pin down any effect. Kruskal-Wallis tests are not feasible because of the small group sizes.

Table 4.17: Boys' computer science performance by group type

Group	Group 1		Group 2		Group 3		Group 4		Group 6	
Variable	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:										
Computer test grade	5	2.90 (0.82)	6	3.08 (0.52)	3	2.58 (0.63)	7	3.00 (0.76)	7	3.07 (0.95)
Comparative performance	5	1.60 (0.89)	6	2.17 (0.75)	3	2.00 (0.00)	7	2.57 (0.98)	7	2.71 (1.25)
Performance measured at end of term:										
Computer test grade	5	3.88 (0.61)	6	4.15 (1.03)	3	1.90 (0.17)	7	3.70 (0.52)	7	3.66 (0.59)
Comparative performance	5	2.40 (0.89)	6	3.00 (0.00)	3	2.33 (1.53)	7	2.71 (0.95)	7	2.57 (0.53)
Expected computer grade	5	2.60 (0.82)	5	2.60 (0.65)	3	2.83 (1.89)	6	1.92 (0.20)	7	2.14 (0.85)
Performance change: +										
Change in test grade	5	0.98 (0.91)	6	1.07 (1.13)	3	-0.68 (0.79)	7	0.70 (0.93)	7	0.59 (0.78)
Change in comparative performance	5	0.80 (0.45)	6	0.83 (0.75)	3	0.33 (1.53)	7	0.14 (1.07)	7	-0.14 (0.90)

Note: Kruskal-Wallis tests are not feasible because of the small sample sizes. Numbers of observations vary due to missing information for some students. + The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students. Results are not shown for group 5 because there are no boys in this group.
Source: Student questionnaires, test results. Own calculations.

In addition to the assessment by group types *via* Kruskal-Wallis tests, the following paragraphs report results from ordinary least squares regressions. The OLS results may be considered to be complementary to the above results. They are presented for the sake of completeness since regression analysis is the tool most often used by empirically oriented economists. Generally, regression analysis provides a simple tool for measuring the effect of interest controlling for different background variables. However, it needs to be verified that the underlying assumptions of the classical linear regression model are met and (in order to use common test statistics and in light of the small sample size) that the errors are normally distributed.

Table 4.18: Simple OLS regressions of mid-term and end of term grades

Outcome variable	Mid-term grade			End of term grade		
Sample	All students	Girls	Girls, Teacher B+	All students	Girls	Girls, Teacher B+
Single-gender effect	0.2972*	0.4267**	0.3308	0.1737	0.3147	-0.2333
(s.e.)	(0.1758)	(0.2046)	(0.2343)	(0.2336)	(0.2856)	(0.3287)
Sample Size	74	46	29	73	45	28

Note: + This sample covers only girls instructed by the same teacher B (who teaches two coeducational and one single-gender groups).
* Significant at the ten percent level. ** Significant at the five percent level.
Source: Student questionnaires, test results. Own calculations.

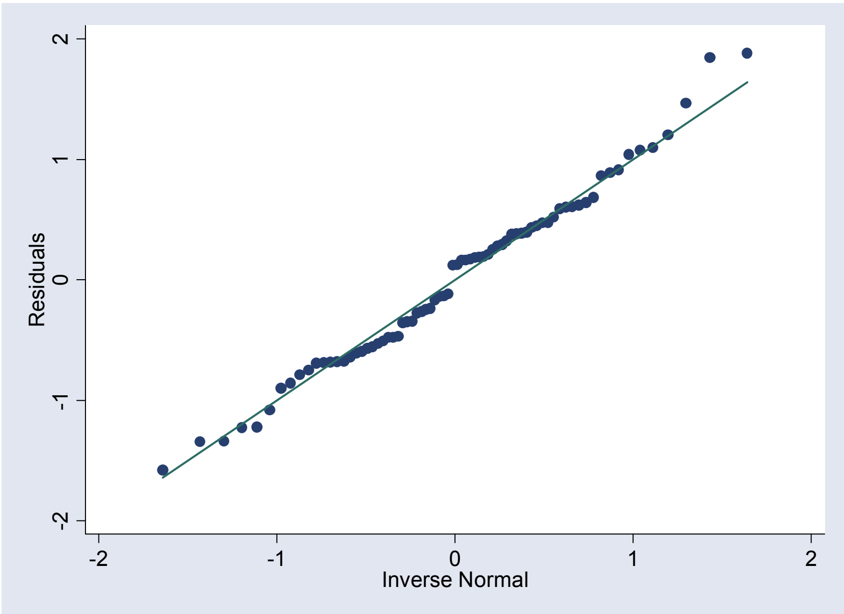
Table 4.18 presents regression results for different samples of interest for the mid-term and end of term grades. Regressing the outcome variables on a dummy variable for single-gender treatment for all students yields a significant effect for the mid-term grade and an insignificant effect for the end of term grade: It seems that students instructed in single-gender classes perform *worse* than students in coeducational classes around mid-term. The effect is especially pronounced if the sample is restricted to girls.[102] However, if the restriction refers to girls taught by the same teacher, the single-gender effect disappears. Therefore, it seems that the effect observed when not considering students taught by the same teachers is an artificial effect due to the teacher and not due to the organizational type.

Further regressions which have been conducted additionally control for students' background variables. However, as one would probably expect (given the limited sample size) the coefficient of the single-gender group indicator is always insignificant. This is why these results are not documented in detail.

[102] By definition, the effects presented here are identical to the ones implied in Table 4.14.

The meaning of the regression results relies (among other assumptions) on the validity of the assumption that the error terms are normally distributed.[103] Figure 4.4 presents the normal probability plot according to Chambers *et al.* (1983), which is a simple check of the assumption that the error terms are normally distributed: Fractiles of the error distribution are plotted versus the fractiles of a normal distribution having the same mean and variance.[104] Eyeballing suggests a rather linear pattern of the normal probability plot, *i.e.* the plot falls close to the diagonal line. Additionally, a Shapiro-Wilk test for normality yields the same result in favour of the normal distribution of the errors. Therefore, the assumption may be taken to be valid.[105]

Figure 4.4: Normal probability plot



Source: Student questionnaires, test results. Own calculations.

Besides the hard performance measures which have been discussed so far, evidence on soft performance measures has been assessed as well. However, in light of the limited sample size it is hard to deduce anything from categorical variables. Therefore, the results are not discussed in detail. Table 4.19 presents the results for girls in the different groups, where grey-shaded columns refer to groups taught by the same teacher.

[103] Several tests have been conducted in order to verify assumptions of the classical linear regression model under which the OLS-estimates have the standard desired properties. Specifically, given the fact that students are taught in different groups, one might assume that errors are heteroskedastic. However, Breusch-Pagan and White-tests for heteroskedasticity based on different sets of variables could not reject the null of homoskedastic errors.
[104] Exemplarily, the presented plot refers to the most general regression discussed above.
[105] The additional test whether the errors are log normally distributed yielded rejection of the null.

Table 4.19: Motivation and gender perception of girls by group

Group	Group 1		Group 2		Group 3		Group 4		Group 5	
Variable	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Variables measures at mid-term:										
I like working with computers.	8	0	5	1	7	2	7	0	15	1
Boys know more about computers.	0	7	0	6	0	9	0	7	0	16
Girls know more about computers.	0	7	2	4	0	9	0	7	6	10
Both genders know about computers	7	0	4	2	9	0	7	0	10	6
I prefer single-gender group.	7	1	3	3	0	9	1	6	7	9
Variables measures at end of term:										
I like working with computers.	4	4	6	0	8	1	5	2	14	2
Boys know more about computers.	1	6	0	6	0	9	0	7	0	16
Girls know more about computers.	1	6	0	6	0	9	2	5	3	13
Both genders know about computers	5	2	6	0	9	0	5	2	13	3
I prefer single-gender group.	4	4	3	3	8	1	2	5	7	9

Source: Student questionnaires. Own calculations.

From the numbers presented in Table 4.19 it is not possible to infer that girls taught in the single-gender group (group 5) differ from girls in the other groups taught by the same teacher (group 2 and 4) concerning the soft performance measures.

4.4 Summary and Discussion of Chapter 4

The present paper documents that there is still plenty of scope for conducting research on the effects of single-gender education. Existing studies are problematic by design since the effects they identify are either only valid for a very specific group or biased because of selection or further implementation problems. Also the presented BW-project does not identify an effect of single-gender education. Performance differences between the single-gender and mixed-gender groups seem to be caused by teacher differences rather than by the organizational treatment. Causal effects are hard to identify in light of a small sample size, confounding factors and sample attrition. Additionally, the small sample based on pupils from one single school's fifth graders may hardly be taken to be representative for the entire German school population. A further drawback of this and many existing studies is that the restriction to a very limited time-span (one observation year) does not allow examining long-run effects.

From all this evidence it is possible to deduce general elements for a potential 'ideal' intervention study. Briefly, in the absence of feasible experiments, features of such a study relate to: (1) large sample sizes together with a standardized assessment of pupils' characteristics in order to be able to control for pupils' selection into different groups using appropriate statistical techniques, (2) a longer project period of several schooling years in order to observe potential long-run effects (and probably also to draw on difference-in-difference techniques), (3) the reduction of confounding factors related to the (educational) treatment of pupils in different groups, *e.g.* through assuring that the same teachers instruct coeducational as well as single-gender groups, (4) the prevention of *Pygmalion* effects, *e.g.* through concealing the gender-related purpose of the intervention project, (5) the reduction of sample attrition (due to missing observations) or group changes through a careful supervision of the project. Especially points (3) to (5) require the support of the participating teachers. Written guidelines and regular meetings of all project partners (teachers and researchers) are certainly important in order to assure the appropriate implementation of the project.

Even if these guidelines are followed, there might still be some problems related to single gender education which have not been addressed so far. Critics of such an intervention study may ask the (legitimate) question why female students should be specifically supported in science studies while male students are not supported in languages and other subjects where they are outperformed by girls. The PISA studies have shown that male students especially, and especially those from the lower secondary schools, show poor reading abilities (cf. for example Röhner, 2003). A stream of the recent education literature emphasizes that most 'problem children' (*e.g.* children displaying aggressive behaviour or other behavioural problems as well as extremely poor school outcomes) are boys and call for new educational measures which focus on boys' development and performance (cf. for example Böhmann, 2003 and Kraus, 1998). In fact, a coherent project could extend the topic of analysis and focus on both genders and different fields of studies.

A further point raised by critics of single gender studies is that it is not clear whether students *want* to be educated separately by gender in the first place (cf. this critique in Biermann, 2000). In the presented BW-project for example, most pupils stated that they prefer to be educated in mixed classes. Related to this issue is the argument that it might be important for girls to learn to succeed in the 'real', competitive and mixed gendered world (cf. Meyer, 1996). On the one hand, single gender education might provide an artificial learning environment and it is not guaranteed that pupils succeeding in this environment will succeed when entering, for example, the mixed gendered labour markets. On the other hand, single gender education might be an appropriate tool for enhancing girl's interest in technical subjects in the first place, while it is possible that these girls are well capable of competing with their male classmates later on (once they have started to put an effort in acquiring the specific skills). Therefore, these arguments tend to support the requirement of conducting *long-term* studies on the effects of single gender education than ruling out the meaning of such studies in the first place.

Researchers opposing single gender experiments often argue that girls' interest and success in computer studies and occupations could be enhanced *within* coeducational classes. Existing studies along this line usually emphasize that the present situation in schools is characterized by co-instruction (instead of coeducation), *i.e.* girls and boys are taught in the same class but a 'secret curriculum' (*heimlicher Lehrplan*) focuses on the abilities and interests of boys. One example is that teachers do not consider different problem-solving strategies of boys and girls but favour the 'male techniques' (cf. Funken *et al.*, 1996, p.128). Concerning different interests related to computer studies, traditional curricula do not consider that girls are less interested in pure programming but more in the application of computer skills, for example using computers for creative tasks (cf. Rentmeister, 1992 and Schultz-Zander, 1992). Further strategies that have been suggested in order to foster girls interest in computers within coeducational classes are for example (1) to discuss the impact of computers on modern societies in the computer class (since girls are usually more interested in social topics) (2) to provide 'good examples', *i.e.* to inform on successful female careers in computer sciences or to (3) increase the number of female computer science teachers.[106] Generally, such approaches which consider the gender perspectives within the coeducational classroom have become known as "reflexive coeducation".

The present (empirical) literature is not conclusive regarding which one of the two ways – single gender education or reflexive coeducation – is more suited to reducing subject-related gender differences. Again, there is plenty of scope for further research on the benefits of single gender education under this perspective.

In summing up, it can be maintained that different streams of discussion call for a more reliable empirical foundation of single gender research. A larger-scale intervention study following the suggested guidelines could crucially improve the insights into this topic given that the presently available evidence is merely speculative.

[106] Cf. Rentmeister, 1992 and, for example, Nossek, 2006 for more specific suggestions of 'reflexive coeducation' in modern computer science classes The study by Hoffmann *et al.* (1997) mentioned above additionally reviews existing evidence and provides evidence on the effects of changed science curricula on girls' and boys' educational outcomes. Lechner (2002) is a follow-up study to the study by Kessels (2002) and emphasizes the role of different educational strategies within science (physics) classes. An extensive discussion and summary of the evidence on reflexive coeducation is given in Faulstich-Wieland (2004b).

References for Chapter 4

Baumert, J. (1992): Koedukation oder Geschlechtertrennung, *Zeitschrift für Pädagogik* 38: 83-110.

Biermann, C. (2000): Lernziel Gleichberechtigung – überflüssiger oder aktueller denn je? Der Versuch einer Zwischenbilanz der Koedukationsdebatte, *Der Deutschunterricht* 52: 90-93.

Billger, Sh.M. (2006): *Reconstructing School Segregation: On the Efficacy and Equity of Single-Sex Schooling*, IZA Discussion Paper No. 2037.

Black, S.E. and A. Spitz-Oener (2007): *Explaining Women's Success: Technological Change and the Skill Content of Women's Work*, NBER Working Paper No. 13116.

BLK (2002): *Frauen in den ingenieur- und naturwissenschaftlichen Studiengängen*, Bericht der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 2.Mai 2002, Bonn.

Böhmann, M. (2003): Jungen in der Schule – ein Problemfall? Die neue Diskussion um die Jungen, *Pädagogik* 55: 32-35.

Brendel, S. (1995): Mädchen in der Hauptschule, in Mack, W. (ed.): *Hauptschule als Jugendschule, Beiträge zur pädagogischen Reform der Hauptschulen in sozialen Brennpunkten*, Süddeutscher Pädagogischer Verlag, Ludwigsburg.

Chambers, J., W. Cleveland, B. Kleiner and P. Tukey (1983): *Graphical Methods for Data Analysis*, Wadsworth International, Duxbury Press, Boston.

Faulstich-Wieland, H. (2004a): Wem nützt die Einrichtung einer Mädchenklasse?, *Zeitschrift für Frauenforschung und Geschlechterstudien* 22: 39-57.

Faulstich-Wieland, H. (2004b): *Mädchen und Naturwissenschaften in der Schule*, Expertise für das Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, Hamburg.

Faulstich-Wieland, H. (1999): Koedukation heute – Bilanz und Chance, in Horstkemperer, M. and M. Kraul (eds.): *Koedukation. Erbe und Chancen*, Deutscher Studien Verlag, Weinheim.

Funken, C., K. Hammerich and B. Schinzel (1996): *Geschlecht, Informatik und Schule*, Academica Verlag, Sankt Augustin.

Häußler, P. and L. Hoffmann (2002): An Intervention Study to Enhance Girls' Interest, Self-Concept, and Achievement in Physics Classes, *Journal of Research in Science Teaching* 39: 870-888.

Häußler, P. and L. Hoffmann (1998): Chancengleichheit für Mädchen im Physikunterricht – Ergebnisse eines erweiterten BLK-Modellversuchs, *Zeitschrift für Didaktik der Naturwissenschaften* 4: 51-67.

Häußler, P. and L. Hoffmann (1990): Physikunterricht für Mädchen und Jungen getrennt? Zwei Interviews aus der Schulpraxis, *Naturwissenschaften im Unterricht. Physik* 1: 32-33.

Heckman, J.J., H. Ichimura and P. Todd (1998): Matching as an Econometric Evaluation Estimator, *Review of Economic Studies* 65: 261-294.

Heckman, J.J. and J.A. Smith (1995): Assessing the Case for Social Experiments, *Journal of Economic Perspectives* 9: 85-110.

Heidtmann, H. (1998): Multimedia und Mädchen, *Praxis Schule 5-10* 9: 25-27.

Heinrichs, U. and T. Schulz (1989): Mädchen und Naturwissenschaften. Wahlverhalten in der gymnasialen Oberstufe, *Pädagogik* 41: 36-39.

Hoffmann, L., P. Häußler and S. Peters-Haft (1997): *An den Interessen von Jungen und Mädchen orientierter Physikunterricht, Ergebnisse eines BLK-Modellversuchs*, Institut für die Pädagogik der Naturwissenschaften, Kiel.

Holz-Ebeling, F., J. Grätz-Tümmers and C. Schwarz (2000): Jungen als „Nutznießer“ der Koedukation? Eine empirische Studie zur Bedeutung der Koedukation für Jungen, *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 32: 94-107.

IW (2005): *Deutschland in Zahlen*, Institut der Deutschen Wirtschaft Köln, Deutscher Instituts-Verlag, Köln.

Kessels, U. (2002): *Undoing Gender in der Schule, Eine empirische Studie über Koedukation und Geschlechtsidentität im Physikunterricht*, Juventa Verlag, Weinheim/München.

Kraus, J. (1998): Koedukation oder Desintegration - Schule auf dem Weg zur Quote? *Profil* 4: 14-15.

Kruskal, W.H. and W.A. Wallis (1952): Use of Ranks in One-Criterion Variance Analysis, *Journal of the American Statistical Association* 47: 583–621.

Lechner, H. (2002): Reflexive Koedukation – eine notwendige Bedingung für den Anfangsunterricht Physik?, in R. Brechel (ed.): *Zur Didaktik der Physik und der Chemie, Tagung 2001*, 86-88.

Ludwig, P.H. (2003): Partielle Geschlechtertrennung – enttäuschte Hoffnung? Monoedukative Lernumgebungen zum Chancenausgleich im Unterricht auf dem Prüfstand, *Zeitschrift für Pädagogik* 49: 640-656.

Mael, F., A. Alonso, D. Gibson, K. Rogers and M. Smith (2005): *Single-Sex Versus Coeducational Schooling: A Systematic Review*, U.S. Department of Education, Policy and Program Study Service Doc. 2005-01, Washington.

Marsh, H.W. and K.J. Rowe (1996): The Effects of Single-Sex and Mixed-Sex Mathematics Classes Within a Coeducational School: A Reanalysis and Comment, *Australian Journal of Education* 40: 147-162.

Marsh, H.W., I.D. Smith, M. Marsh and L. Owens (1988): The Transition From Single-Sex to Coeducational High Schools: Effects on Multiple Dimensions of Self-Concept and on Academic Achievement, *American Educational Research Journal* 25: 237-269.

Meyer, M. (1996): Wenn die Mädchen dabei gewesen wären… Schülerinnen und Schüler äußern sich über ITG-Unterricht in geschlechtshomogenen Gruppen, *Computer und Unterricht* 6: 50-52.

Nossek, G. (2006): Reflexive Koedukation, *Computer und Unterricht* 16: 36-37.

Nyssen, E. (1996): *Mädchenförderung in der Schule, Ergebnisse und Erfahrungen aus einem Modellversuch*, Juventa Verlag, Weinheim/München.

Rastatt (2005): *Rastatt in Zahlen*, Stadt Rastatt.

Rentmeister, C. (1992): Computer und Kreativität, in Glumper, E. (eds.): *Mädchenbildung – Frauenbildung*, Verlag Julius Klinkhard, Bad Heilbrunn, 221-227.

Roeder, P. and S. Gruehn (1997): Geschlecht und Kurswahlverhalten, *Zeitschrift für Pädagogik* 43: 877-894.

Röhner, C. (2003): Bildungsverlierer: Jungen? Zur Koedukationsdebatte nach PISA, *Praxis Schule 5-10* 14: 11-15.

Rost, D.H. and C. Pruisken (2000): Vereint schwach? Getrennt stark? Mädchen und Koedukation, *Zeitschrift für pädagogische Psychologie* 14: 177-193.

Rustmeyer, R. and A. Jubel (1996): Geschlechtsspezifische Unterschiede im Unterrichtsfach Mathematik hinsichtlich der Fehleinschätzung, Leistungserwartung, Attribution sowie im Lernaufwand und im Interesse, *Zeitschrift für Pädagogische Psychologie* 10: 13-25.

Schuld, E. (1997): Das Internet: Benachteiligung oder Förderung von Mädchen am Computer?, *Didaktik der Berufs- und Arbeitswelt* 4: 21-30.

Schultz-Zander, R. (1992): Für die Gleichstellung von Mädchen und jungen Frauen in der Informationstechnischen Bildung, in E. Glumper (ed.): *Mädchenbildung – Frauenbildung*, Verlag Julius Klinkhard, Bad Heilbrunn, 210-220.

Thierack, A. (1995): In der Schule benachteiligt, in der Koedukationsdebatte vergessen: Lebenswelten von HauptschülerInnen, in M. Hempel (ed.): *Verschieden und doch gleich. Schule und Geschlechterverhältnisse in Ost und West*, Klinkhardt, Bad Heilbrunn.

Volmerg, B., A. Creutz, M. Reinhardt and T. Eiselen (1996): *Ohne Jungs ganz anders? Geschlechterdifferenz und Lehrerrolle am Beispiel eines Schulversuchs*, Kleine Verlag, Bielefeld.

Ziegler, A., P. Broome and K.A. Heller (1998): Pygmalion im Mädchenkopf, Erwartungs- und Erfahrungseffekte koedukativen vs. geschlechtshomogenen Physikanfangsunterrichts, *Psychologie in Erziehung und Unterricht* 45: 2-18.

Appendix 4A: Appendix of Questionnaires

The following pages display the original pupil, parent and teacher questionnaires (in German language) distributed in the first year of the BW-project. The same questions are repeated in the second-year questionnaires. Appendix 4C provides descriptive statistics for the variables drawn from these questionnaires.

Schulprojekt: Computer in der Schule

**Schülerfragebogen (wird in der Schule ausgefüllt)
zum Halbjahr, Klasse 5a Gruppe 1**

Bitte kreuze nur die Antworten an, die auf dich zutreffen: ☒

1. Bitte gib die Nummer an, die dir dein Lehrer für diese Befragung gegeben hat:

☐

2. Wie alt bist du?

☐ Jahre

3. Hat deine Familie zu Hause einen Computer?

Ja ☐ Nein ☐

4. Wie viel Spaß macht dir der Unterricht am Computer? Kreuze bitte das an, was zutrifft.

Sehr viel Spaß ____ ☐
Eher viel Spaß ____ ☐
Weder viel noch wenig Spaß______________________________ ☐
Eher wenig Spaß ______________________________________ ☐
Keinen Spaß ____ ☐

5. Wie sind bisher in der Hauptschule deine Leistungen allgemein im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere. ________________ ☐
Ich erbringe bessere Leistungen als andere. ____________________ ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere.__ ☐
Andere erbringen bessere Leistungen. _________________________ ☐
Andere sind viel besser.___________________________________ ☐

6. Wie sind bisher in der Hauptschule deine Leistungen *im Computerunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere. ________________ ☐
Ich erbringe bessere Leistungen als andere. ____________________ ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere.__ ☐
Andere erbringen bessere Leistungen. _________________________ ☐
Andere sind viel besser.___________________________________ ☐

- Fragebogen: Schüler zum Halbjahr, Seite 1 -

7. Wie sind bisher in der Hauptschule deine Leistungen *im Matheunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

- Ich erbringe viel bessere Leistungen als andere. ☐
- Ich erbringe bessere Leistungen als andere. ☐
- Ich erbringe weder bessere noch schlechtere Leistungen als andere. ☐
- Andere erbringen bessere Leistungen. ☐
- Andere sind viel besser. ☐

8. Wie sind bisher in der Hauptschule deine Leistungen *im Deutschunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

- Ich erbringe viel bessere Leistungen als andere. ☐
- Ich erbringe bessere Leistungen als andere. ☐
- Ich erbringe weder bessere noch schlechtere Leistungen als andere. ☐
- Andere erbringen bessere Leistungen. ☐
- Andere sind viel besser. ☐

9. Was würdest du am Unterricht am Computer verbessern?

____

____

____

____

10. Beantworte diese Frage nur, wenn du ein Mädchen bist: Findest du es angenehmer, in einer Klasse nur mit Mädchen zu sein?
Kreuze bitte an, was du gut findest!

Nur mit Mädchen ☐ Mit Jungen und Mädchen ☐

11. Beantworte diese Frage nur, wenn du ein Junge bist: Findest du es angenehmer, in einer Klasse nur mit Jungen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Jungen ☐ Mit Jungen und Mädchen ☐

12. Wer meinst du, kennt sich besser mit Computern aus: eher die Jungs oder eher die Mädchen?

Eher die Jungs ☐ Eher die Mädchen ☐ Beide ☐

Schulprojekt: Computer in der Schule

Schülerfragebogen (wird in der Schule ausgefüllt)
zum Schuljahresende, Klasse 5a, Gruppe 1

Bitte kreuze nur die Antworten an, die auf dich zutreffen: ☒

1. Bitte gib die Nummer an, die dir dein Lehrer für diese Befragung gegeben hat:

[ ]

2. Hat deine Familie zu Hause einen Computer? Ja ☐ Nein ☐

3. Wie viel Spaß macht dir der Unterricht am Computer? Kreuze bitte das an, was zutrifft.

   Sehr viel Spaß............ ☐
   Eher viel Spaß............ ☐
   Weder viel noch wenig Spaß............ ☐
   Eher wenig Spaß............ ☐
   Keinen Spaß............ ☐

4. Wie sind bisher in der Hauptschule deine Leistungen allgemein im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere............ ☐
   Ich erbringe bessere Leistungen als andere............ ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere... ☐
   Andere erbringen bessere Leistungen............ ☐
   Andere sind viel besser............ ☐

5. Wie sind bisher in der Hauptschule deine Leistungen *im Computerunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere............ ☐
   Ich erbringe bessere Leistungen als andere............ ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere... ☐
   Andere erbringen bessere Leistungen............ ☐
   Andere sind viel besser............ ☐

6. Wie sind bisher in der Hauptschule deine Leistungen *im Matheunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere............ ☐
   Ich erbringe bessere Leistungen als andere............ ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere... ☐
   Andere erbringen bessere Leistungen............ ☐
   Andere sind viel besser............ ☐

7. Wie sind bisher in der Hauptschule deine Leistungen *im Deutschunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere............ ☐
   Ich erbringe bessere Leistungen als andere............ ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere... ☐
   Andere erbringen bessere Leistungen............ ☐
   Andere sind viel besser............ ☐

- Fragebogen: Schüler zum Halbjahr, Seite 1 -

8. Gib eine Gesamtnote zwischen 1 und 6 an, die Deine Leistungen beschreibt, die Du durchschnittlich im Zeugnis erwartest: ☐

9. Gib eine Gesamtnote zwischen 1 und 6 an, die Deine Leistungen beschreibt, die Du im Computerunterricht im Zeugnis erwartest: ☐

10. Welche Note bekommst Du wahrscheinlich in Deutsch? ☐

11. Welche Note bekommst Du wahrscheinlich in Mathe? ☐

12. Wie viele Brüder hast Du?

☐ Brüder

13. Wie viele Schwestern hast Du?

☐ Schwestern

14. Wie viele Kinder (ohne Dich) leben bei Dir zu Hause?

☐ Kinder

15. Wie alt sind diese Kinder, die bei Dir zu Hause wohnen?

Alter des 1. Kindes (falls vorhanden): ☐ Jahre
Alter des 2. Kindes (falls vorhanden): ☐ Jahre
Alter des 3. Kindes (falls vorhanden): ☐ Jahre
Alter des 4. Kindes (falls vorhanden): ☐ Jahre
Alter des 5. Kindes (falls vorhanden): ☐ Jahre
Alter des 6. Kindes (falls vorhanden): ☐ Jahre

16. Wurdest Du in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

17. Wurde Dein Vater in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

18. Wurde Deine Mutter in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

19. Sprecht Ihr zu Hause eine Sprache, die nicht die deutsche Sprache ist?

Ja.. ☐ Nein.. ☐

- Fragebogen: Schüler zum Halbjahr, Seite 2 -

20. Wenn Ihr zu Hause eine Sprache sprecht, die nicht die deutsche Sprache ist, welche Sprache ist das? Schreibe die Sprache in das Kästchen.

☐

21. Wohnen Deine Großeltern (z.B. Deine Oma oder Dein Opa) bei Dir zu Hause oder in der Nähe?

Ja..☐ Nein.. ☐

22. Welchen Schulabschluss hat Deine Mutter? Kreuze das Zutreffende an!

Keinen..☐
Hauptschule (oder vergleichbar)..☐
Realschule (oder vergleichbar)...☐
Gymnasium (oder vergleichbar)..☐
Weiß ich nicht.. ☐

23. Welchen Schulabschluss hat Dein Vater? Kreuze das Zutreffende an!

Keinen..☐
Hauptschule (oder vergleichbar)..☐
Realschule (oder vergleichbar)...☐
Gymnasium (oder vergleichbar)..☐
Weiß ich nicht.. ☐

24. Hast Du einen Vorschlag, wie der Unterricht am Computer besser werden könnte?

---

25. Beantworte diese Frage nur, wenn du ein Mädchen bist: Findest du es angenehmer, im Computerunterricht in einer Klasse nur mit Mädchen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Mädchen....☐ Mit Jungen und Mädchen....☐

26. Beantworte diese Frage nur, wenn du ein Junge bist: Findest du es angenehmer, im Computerunterricht in einer Klasse nur mit Jungen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Mädchen....☐ Mit Jungen und Mädchen....☐

27. Wer meinst du, kennt sich besser mit Computern aus: eher die Jungs oder eher die Mädchen?

Eher die Jungs...☐ Eher die Mädchen...☐ Beide...☐

Schüler- Nr.:

Schulprojekt: Computer in der Schule

Fragebogen: Eltern

Bitte kreuzen Sie nur die Antworten an, die auf Sie zutreffen: ☒

1. Sind Sie die Mutter (bzw. weibliche Bezugsperson) oder der Vater (bzw. die männliche Bezugsperson) des Kindes in der 5. Klasse?

   Mutter (bzw. weibliche Bezugsperson): ☐
   Vater (bzw. männliche Bezugsperson): ☐

2. Sind Sie allein erziehend?

   Ja ☐ Nein ☐

3. Bitte geben Sie die Gesamtzahl Ihrer Kinder an: ☐

4. Wie viele Kinder leben in Ihrem Haushalt? ☐

5. Wie alt sind diese Kinder? Bitte geben Sie das Alter aller Kinder, die in Ihrem Haushalt leben, in Jahren an (z.B. schreiben Sie die Zahl 8 in ein Kästchen, wenn Ihr Kind 8 Jahre alt ist).

   Alter des ersten Kindes: ☐ Jahre
   Alter des zweiten Kindes (falls vorhanden): ☐ Jahre
   Alter des dritten Kindes (falls vorhanden): ☐ Jahre
   Alter des vierten Kindes (falls vorhanden): ☐ Jahre
   Alter des fünften Kindes (falls vorhanden): ☐ Jahre

6. Wohnt ein Teil der Großeltern (ein Opa oder eine Oma) des Kindes in Ihrem Haushalt oder in der Nähe?

   Ja ☐ Nein ☐

7. Wurde Ihr Kind, welches die 5. Klasse besucht, in Deutschland geboren?

   Ja ☐ Nein ☐

8. Wurden Sie in Deutschland geboren?

   Ja ☐ Nein ☐

9. Wurde der andere Elternteil des Kindes beziehungsweise die andere erwachsene Bezugsperson des Kindes in Deutschland geboren?

   Ja ☐ Nein ☐

10. Welche Staatsbürgerschaft besitzen Sie (z.B. Deutsch)? ☐

11. Welche Staatsbürgerschaft besitzt die andere erwachsene Bezugsperson des Kindes? [ ]

12. Sprechen Sie mit Ihrem Kind eine Sprache, die nicht die deutsche Sprache ist?

Ja ☐ Nein ☐

13. Wenn Sie Frage 12 mit Ja beantwortet haben: Welche Sprache ist das?

[ ]

14. Haben Sie in Ihrem Beruf mit Computern zu tun?

Ja ☐ Nein ☐

15. Wenn Sie Frage 14 mit Ja beantwortet haben: Beschreiben Sie, auf welche Art und Weise Sie in Ihrem Beruf Computer nutzen:

____

____

____

____

16. In welchem Umfang sind Sie berufstätig?

Nicht berufstätig ☐ Geringfügig beschäftigt ☐
Teilzeit beschäftigt ☐ Vollzeit beschäftigt ☐

17. In welchem Umfang ist der andere Elternteil des Kindes (beziehungsweise die andere erwachsene Bezugsperson des Kindes) berufstätig?

Nicht berufstätig ☐ Geringfügig beschäftigt ☐
Teilzeit beschäftigt ☐ Vollzeit beschäftigt ☐

18. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) beruflich mit Computern zu tun?

Ja ☐ Nein ☐

19. Wenn Sie Frage 18 mit Ja beantwortet haben: Beschreiben Sie, auf welche Art und Weise der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) beruflich einen Computer nutzt:

____

____

- Fragebogen: Eltern, Seite 2 -

____

____

20. Welchen allgemeinen Schulabschluss (Hauptschule, Realschule oder Gymnasium) besitzen Sie?

Keinen allgemeinen Schulabschluss..☐

Hauptschulabschluss
oder vergleichbaren Abschluss, z.B. im Ausland erworben ....☐

Mittlere Reife / Realschulabschluss oder vergleichbaren Abschluss....☐

Abitur, Fachhochschulreife od. vergleichbaren Abschluss ....☐

21. Welchen allgemeinen Schulabschluss (Hauptschule, Realschule oder Gymnasium) besitzt der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes)?

Keinen allgemeinen Schulabschluss..☐

Hauptschulabschluss
oder vergleichbaren Abschluss, z.B. im Ausland erworben ....☐

Mittlere Reife / Realschulabschluss oder vergleichbaren Abschluss....☐

Abitur, Fachhochschulreife od. vergleichbaren Abschluss ....☐

22. Haben Sie einen beruflichen Abschluss (z.B. Lehre)?

Ja ☐ Nein ☐

23. Wenn Sie Frage 22 mit Ja beantwortet haben: Welchen beruflichen Abschluss besitzen Sie?

24. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) einen beruflichen Abschluss (z.B. Lehre)?

Ja ☐ Nein ☐

25. Wenn Sie Frage 24 mit Ja beantwortet haben: Welcher Abschluss ist das?

26. Haben Sie studiert?

Ja ☐ Nein ☐

- Fragebogen: Eltern, Seite 3 -

27. Wenn Sie Frage 26 mit Ja beantwortet haben: Welchen (Fach-) Hochschulabschluss besitzen Sie? ☐

28. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) studiert?

Ja ☐ Nein ☐

29. Wenn Sie Frage 28 mit Ja beantwortet haben: Welchen Hochschulabschluss besitzt die Person? ☐

30. Welches Bruttoeinkommen (einschließlich Einkommen des im Haushalt lebenden Partners, Sozialhilfe etc.) steht Ihrem Haushalt monatlich zur Verfügung? Kreuzen Sie die relevante Einkommensklasse an!

weniger als 800 €	☐	800-1200 €	☐
1200-1500 €	☐	1500-1800 €	☐
1800-2200 €	☐	2200-2500 €	☐
2500-2800 €	☐	2800-3200 €	☐
3200-3700 €	☐	mehr als 3700 €	☐

31. Wurde im Grundschulunterricht Ihres Kindes, welches in der 5. Klasse ist, ab und zu ein Computer eingesetzt?

Ja ☐ Nein ☐

32 Wie wichtig finden Sie, dass Ihr Kind ab Klasse 5 Computerunterricht erhält?

sehr wichtig ☐, wichtig ☐, weniger wichtig ☐, nicht wichtig ☐

33. Haben Sie selbst einen Computer bei der Hausaufgabenvorbereitung Ihres Kindes, welches in der 5. Klasse ist, eingesetzt?

Ja ☐ Nein ☐

34. Wie viele Stunden verbringt Ihr Kind an einem durchschnittlichen Tag in seiner Freizeit am Computer?

☐ Stunden

35. Welche Note hatte Ihr Kind im letzten Zeugnis in Mathematik? ☐

36. Welche Note hatte Ihr Kind im letzten Zeugnis in Deutsch? ☐

37. Vergeben Sie eine Gesamtnote zwischen 1 und 6, die die Leistung Ihres Kindes in der Grundschule beschreibt: ☐

Vielen Dank für Ihre Teilnahme!

Schulprojekt: Computer in der Schule

Lehrerfragebogen
Klasse 5a, Gruppe 1

A Fragen zur Gruppenstruktur

1. Wie viele Kinder ausländischer Abstammung (d.h. im Ausland geboren oder Kinder, deren Eltern im Ausland geboren sind) sind etwa in dieser Gruppe?

☐ Schüler

2. Wie viele Schüler haben **sprachliche** Probleme, dem Unterricht zu folgen?

☐ Schüler

3. Gibt es in der Gruppe Probleme bezüglich des Verhaltens / des Zusammenhalts der Kinder untereinander oder besteht ein gutes Verhältnis unter den Kindern?

Sehr gutes Verhältnis der Kinder in der Gruppe zueinander............................ ☐
Eher gutes Verhältnis der Kinder in der Gruppe zueinander........................... ☐
Weder besonders gutes noch besonders schlechtes Verhältnis der Kinder..... ☐
Eher schlechtes Verhältnis der Kinder in der Gruppe zueinander.................. ☐
Problematisches Verhältnis der Kinder in der Gruppe zueinander................. ☐

4. Gab es in diesem Schuljahr Fluktuation in der Gruppe, d.h. sind während des Schuljahres Kinder abgegangen oder neu in die Schule gekommen?

Ja ☐ Nein ☐

B Fragen zur Gruppenmotivation

1. Wie motiviert schätzen Sie die Schüler dieser Gruppe durchschnittlich ein? Bitte beziehen Sie sich bei dieser Aussage auf den Computerunterricht.

Sehr motiviert.. ☐
Eher motiviert.. ☐
Weder motiviert noch unmotiviert.. ☐
Eher wenig motiviert.. ☐
Unmotiviert.. ☐

2. Gibt es in dieser Gruppe Schüler, die Sie als Problemschüler bezeichnen würden?

Ja ☐ Nein ☐

3. Wenn Sie Frage 2 mit „Ja“ beantwortet haben, auf wie viele Schüler trifft das zu?

☐ Schüler

4. Bitte beschreiben Sie ggf., welche Probleme mit den Schülern bestehen.

- Fragebogen: Schüler zum Halbjahr, Seite 1 -

5. Gibt es häufige Störungen des Unterrichts?

Ja ☐ Nein ☐

6. Wenn Sie Frage 5 mit „Ja“ beantwortet haben, welche Störungen sind das?

7. Wie viele Schüler würden Sie als **inhaltlich** mit dem Unterrichtsstoff als überfordert bezeichnen?

☐ Schüler

8. Welchen Notendurchschnitt erwarten Sie ungefähr für diese Gruppe in diesem Unterricht am Ende des Schuljahres?

☐

9. Wie häufig müssen die Schüler für den Computerunterricht Hausaufgaben machen?

Jede Woche.. ☐
Alle zwei Wochen... ☐
Etwa ein- bis zweimal im Monat.......................... ☐
Weniger als einmal im Monat.............................. ☐
Nie... ☐

Zum Schluss habe ich noch einige Fragen zu Ihrer Person:

1. Unterrichten Sie diese Klasse auch in anderen Fächern?

Ja ☐ Nein ☐

2. Wenn Sie Frage mit „Ja“ beantwortet haben, welche Fächer sind das?

3. Welche Fächer unterrichten Sie allgemein?

- Fragebogen: Schüler zum Halbjahr, Seite 2 -

4. Wie finden Sie die Computerausstattung dieser Schule: Ist diese für die Bedürfnisse des Unterrichts ausreichend?

Ja ☐ Nein ☐

5. Wie gut klappt die Zusammenarbeit mit den Schülerinnen und Schülern in dieser Gruppe allgemein?

Sehr gut... ☐
Gut... ☐
Weder besonders gut noch schlecht... ☐
Eher problematisch.. ☐
Sehr problematisch... ☐

6. Was ist Ihre Meinung: Welche Kinder sind begabter im Computerunterricht?

Mädchen...☐ Jungen...☐ kann man so nicht sagen...☐

7. Was ist Ihre Meinung: Ist es sinnvoll, Jungen und Mädchen im Computerunterricht getrennt zu unterrichten?

Ja...☐ Nein...☐ weiß ich nicht...☐

8. Warum? (bezieht sich auf die Antwort zu Frage 7)

---

9. Aus Ihrer Berufserfahrung heraus, welchen der folgenden Aussagen würden Sie zustimmen?

Mädchen sind eher sprachlich begabt als Jungen:
stimmt...☐ stimmt nicht...☐

Jungen sind eher begabt für Mathematik oder Naturwissenschaften als Mädchen:
stimmt...☐ stimmt nicht...☐

Unterschiedliche Stärken und Schwächen von Jungen und Mädchen für verschiedene Fächer sind wahrscheinlich eher durch die Erziehung bedingt als natürlich gegeben:
stimmt...☐ stimmt nicht...☐

Unterschiedliche Stärken und Schwächen von Jungen und Mädchen für verschiedene Fächer sind wahrscheinlich sowohl durch die Erziehung als auch durch natürliche Veranlagung bestimmt:
stimmt...☐ stimmt nicht...☐

Jungen sollten eher technische Berufe anstreben als Mädchen:
stimmt...☐ stimmt nicht...☐

- Fragebogen: Schüler zum Halbjahr, Seite 3 -

Mädchen sind nicht begabt für technische Berufe:

stimmt...☐ stimmt nicht...☐

Mädchen haben im Durchschnitt stärkere Probleme, dem Unterrichtsstoff der Hauptschule zu folgen als Jungen:

stimmt...☐ stimmt nicht...☐

Jungen haben im Durchschnitt stärkere Probleme, dem Unterrichtsstoff der Hauptschule zu folgen als Mädchen:

stimmt...☐ stimmt nicht...☐

10. Wie lange üben Sie diesen Beruf aus?

☐ Jahre

11. Wie alt sind Sie?

☐ Jahre

12. Haben Sie sonstige Anmerkungen zum Unterricht mit dieser Gruppe oder zu dieser Umfrage?

___

___

___

___

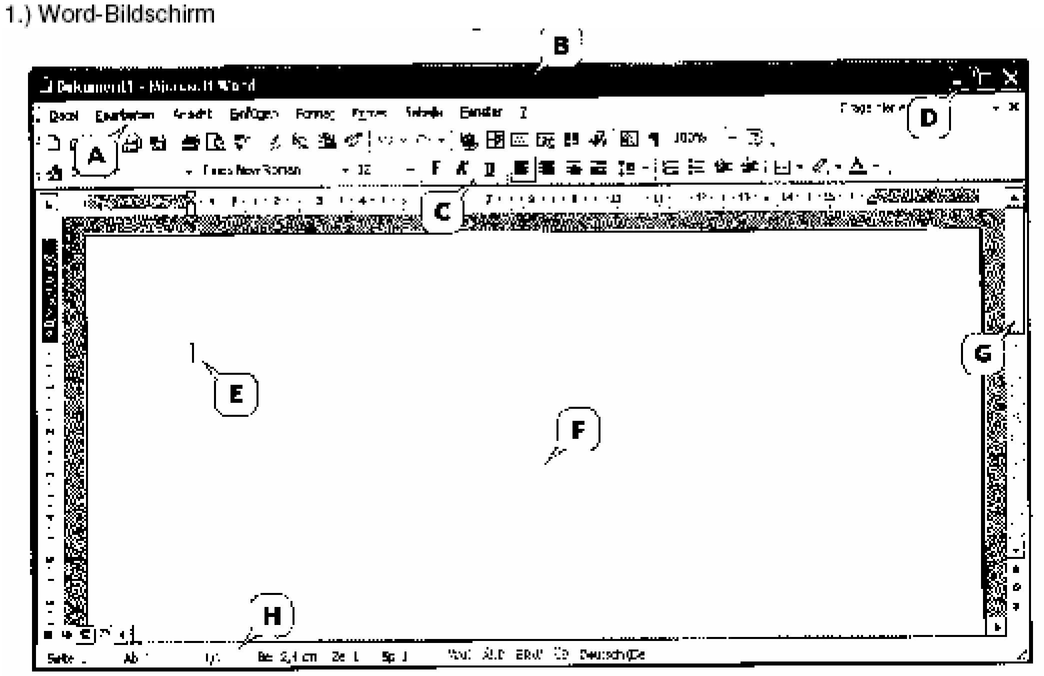
Vielen Dank für Ihre Teilnahme!

Appendix 4B: Appendix of Tests

The following pages display the original tests (in German language) that were conducted in the first year of the BW-project. English translations of the test questions are given below the tests.

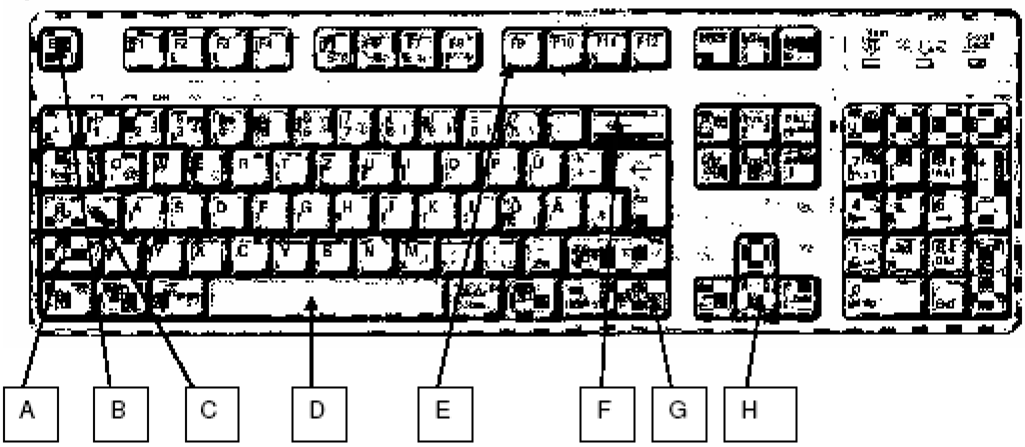
Arbeit Informatik Kl. 5

1.) Word-Bildschirm



Ordne folgende Begriff den Buchstaben zu!
Menüleiste Cursor Titelleiste Statuszeile Symbolleiste Bildlaufleiste Arbeitsfeld Schaltfläche in der Titelleiste 4P

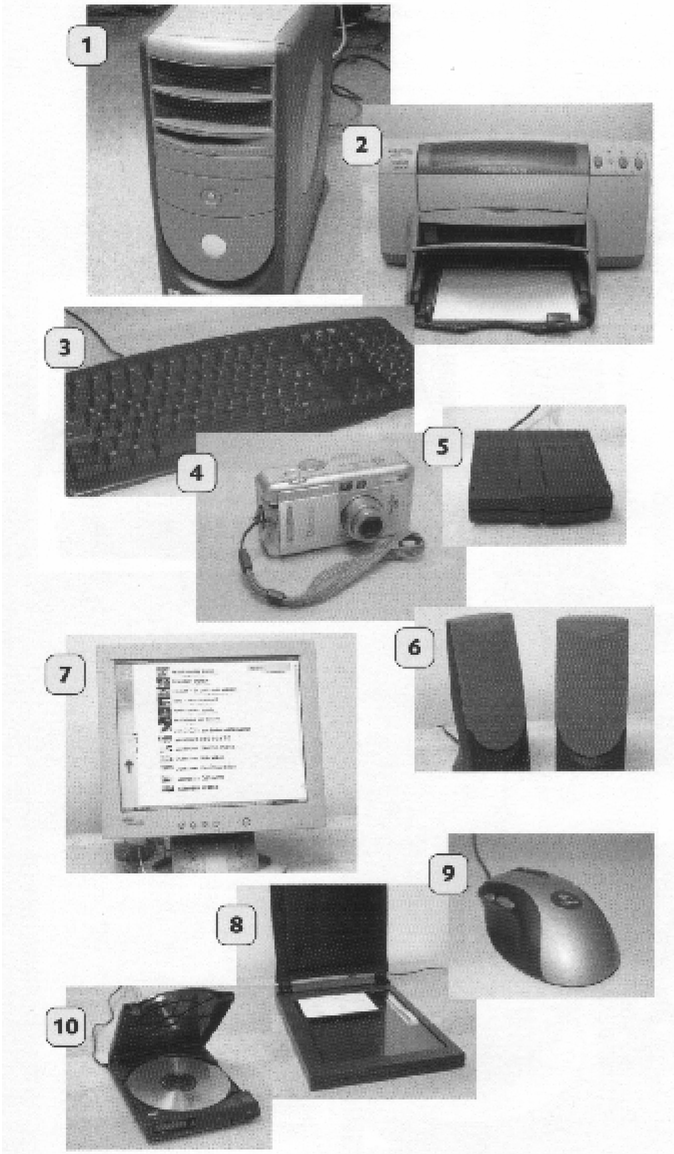
2.) Tastatur



Benenne die Tasten!

A= B=
C= D=
E= F=
G= H= 4P

3.) Teile einer Computer-Anlage



1= ____________________

2= ____________________

3= ____________________

4= ____________________

5= ____________________

6= ____________________

7= ____________________

8= ____________________

9= ____________________

10= ____________________

5P

gesamt 13P

English translation of the test questions (first test):

1) Word-Screen: Which letters correspond to the following terms: menu bar, cursor, title bar, status bar, toolbar, scrollbar, workplace, push button in the title bar?
2) Keyboard: Name the keys!
3) Name the parts of the computer equipment!

Informatik-Test Kl.5 Datum:____________

Name:____________________________ Klasse:_____

1. Beschreibe wie du deine fertigen Dateien speicherst!

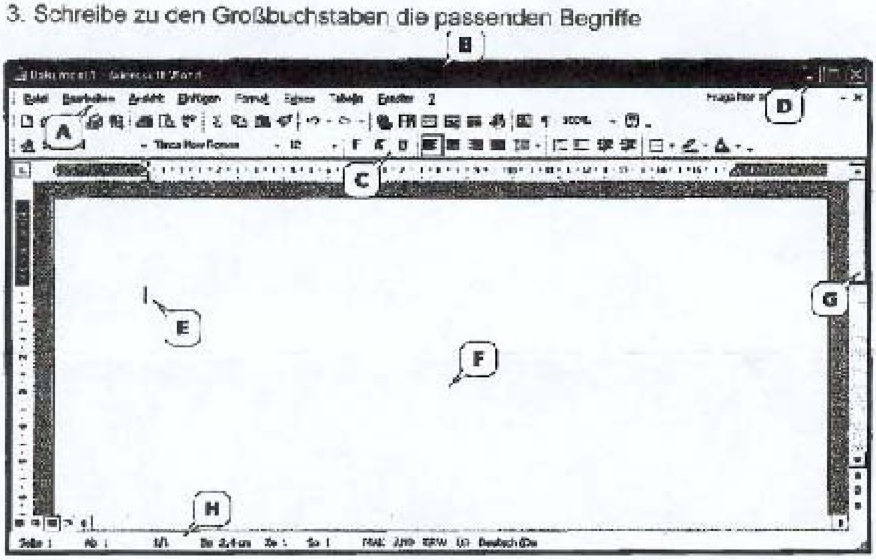
____

____

2. Welchen Arbeitsschritt musst du beachten bevor du einen Text ausdruckst?

____

3. Schreibe zu den Großbuchstaben die passenden Begriffe



A:__________________________ B:__________________

C:__________________________, D __________________

E:__________________________, F:__________________

G:__________________________, H:__________________

4. Beschreibe in Stichworten, wie das Internet funktioniert!

____

____

5. Nenne die Teile, die zu einer Computeranlage gehören!

____

____

____

English translation of the test questions (second test):

1) Explain how to save a file!

2) What do you need to consider before printing a text file?

3) Which terms correspond to the indicated capital letters?

4) How does the internet work (provide some keywords)?

5) What are the parts of the computer equipment?

Appendix 4C: Appendix on the Students' Background

Table 4C.1 in this Appendix shows descriptive statistics for major variables describing the general background of the students considered (fifth-graders in 2004/2005). The information is drawn from the students' questionnaires which were distributed in January and June 2005.[107] Generally, girls are a majority in this cohort of lower secondary school students (62 %). The variation in the students' age is rather high, with the youngest students being ten and a half years old and the oldest being more than thirteen years old in January 2005. The average student is eleven and a half years old.

Even if most students were born in Germany (77 %), the majority of them have some immigrant background (in the sense that at least one parent was born abroad or a foreign language is spoken at home). The languages spoken at home, which are reported in the student questionnaires, indicate that most immigrants originate from Russia (42 %) and Turkey (24 %). Fewer are immigrants from the former Yugoslavia (9 %) and Italy (9 %) or Romania (7 %). Three remaining students seem to have an immigrant background from Hungary, Thailand and some Arabian country, respectively.[108]

The average student has one or two siblings. The majority of siblings (36 % of all siblings) are more than three years younger than the considered student. Concerning parental background, parental education is unfortunately missing in many cases: 19 % of the students do not know about their mother's secondary education and 27 % are not able to indicate their father's secondary education. Among those who reported parental education, 39% (18 %) state that their mother's highest secondary degree corresponds to a low (an intermediate) degree, while fewer mothers seem to hold no secondary degree (12 %) or a higher level secondary degree (12 %). Similarly, among the fathers, most reported degrees refer to the lower secondary schools (35 %); only 20 % and 12 % hold intermediate and higher degrees respectively and 5 % seem to hold no secondary degree at all. These numbers indicate that the parents of lower-level high school students generally tend to hold lower (or, at most, intermediate) degrees as well.

Table 4C1 includes two variables which can be considered as proxies for 'family care': First of all there is a dummy variable referring to a grandparent living in the same house or close. Nearly half of the students report that a grandparent is living nearby. Secondly, parental interest is proxied assuming that parents are at least to some extent interested in the child's school performance if they opted to complete the voluntary parents' question-

[107] In few cases this information is combined with data from the parents' questionnaires because the reported answers were not clear or did not seem reliable.
[108] This information is confirmed by information from the parents' questionnaires.

naire: 66 % of the children seem to have a mother or father who cares about school matters and completed the questionnaire.

Table 4C.1: Descriptive statistics for major background characteristics

Variable	Means	(s.d.)	Min	Max
Gender indicator (male = 1)	0.38	(0.49)	0.00	1.00
Age of students (in January 2005)	11.53	(0.61)	10.52	13.13
Immigrant background (Reference: no such background):				
Student: born in Germany	0.77	(0.42)	0.00	1.00
Both parents born in Germany	0.30	(0.46)	0.00	1.00
Only German language used at home	0.35	(0.48)	0.00	1.00
Some immigrant background	0.78	(0.41)	0.00	1.00
Number of siblings …				
…absolute	1.54	(1.06)	0.00	5.00
…more than 3 years younger	0.55	(0.83)	0.00	4.00
…3 or less than 3 years younger	0.34	(0.56)	0.00	2.00
…older, but at most 3 years	0.19	(0.39)	0.00	1.00
…more than 3 years older	0.36	(0.61)	0.00	2.00
Parental education dummies				
Mother’s education: missing value	0.19	(0.39)	0.00	1.00
Mother’s education: no secondary degree	0.12	(0.33)	0.00	1.00
Mother’s education: lower secondary degree	0.39	(0.49)	0.00	1.00
Mother’s education: intermediate degree	0.18	(0.38)	0.00	1.00
Mother’s education: higher secondary degree	0.12	(0.33)	0.00	1.00
Father’s education: missing value	0.27	(0.45)	0.00	1.00
Father’s education: no secondary degree	0.05	(0.23)	0.00	1.00
Father’s education: lower secondary degree	0.35	(0.48)	0.00	1.00
Father’s education: intermediate degree	0.20	(0.40)	0.00	1.00
Father’s education: higher secondary degree	0.12	(0.33)	0.00	1.00
Family Care:				
Grandparents live nearby (dummy variable)	0.46	(0.50)	0.00	1.00
Parental interest (participation in project dummy)	0.66	(0.48)	0.00	1.00
Number of observations		74		

Note: The calculations only include students observed throughout the year since most background characteristics are only available for these students.
Source: Student questionnaires. Own calculations.

Table 4C.2 of this Appendix reports means of the considered background variables separately by gender. Most variables take a similar value for boys and girls. However, it is obvious that girls more often fail to report their fathers’ secondary degree (30 %) than boys (21 %), while boys more often state that they do not know about their mothers’ secondary education (25 %) than girls (15 %). In addition, the further parental education categories seem to differ according to the gender of the student (for those who reported on parental education). For example, 20 % of the girls declare that their father holds a higher secondary degree, while none of the boys’ fathers seems to hold such a degree. It is hard to tell if this is due to an incorrect assignment of degrees by the children or to real differences in the parents’ education.

Table 4C.2: Descriptive statistics for major background characteristics by gender

Variable / Group	Girls		Boys	
	Means	(s.d.)	Means	(s.d.)
Age of students (in January 2005)	11.44	0.58	11.67	0.64
Immigrant background (Reference: no such background):				
Student: born in Germany	0.74	0.44	0.82	0.39
Both parents born in Germany	0.33	0.47	0.25	0.44
Only German language used at home	0.37	0.49	0.32	0.48
Some immigrant background*	0.76	0.43	0.82	0.39
Number of siblings …				
…absolute	1.65	1.10	1.36	0.99
…more than 3 years younger	0.57	0.83	0.54	0.84
…3 or less than 3 years younger	0.35	0.57	0.32	0.55
…older, but at most 3 years	0.22	0.42	0.14	0.36
…more than 3 years older	0.43	0.69	0.25	0.44
Parental education dummies				
Mother's education: missing value	0.15	0.36	0.25	0.44
Mother's education: no secondary degree	0.17	0.38	0.04	0.19
Mother's education: lower secondary degree	0.39	0.49	0.39	0.50
Mother's education: intermediate degree	0.13	0.34	0.25	0.44
Mother's education: higher secondary degree	0.15	0.36	0.07	0.26
Father's education: missing value	0.30	0.47	0.21	0.42
Father's education: no secondary degree	0.04	0.21	0.07	0.26
Father's education: lower secondary degree	0.30	0.47	0.43	0.50
Father's education: intermediate degree	0.15	0.36	0.29	0.46
Father's education: higher secondary degree	0.20	0.40	0.00	0.00
Family Care:				
Grandparents live nearby (dummy variable)	0.48	0.51	0.43	0.50
Parental interest (participation in project dummy)	0.65	0.48	0.68	0.48
Number of observations	46		28	

Note: The calculations only include students observed throughout the year since most background characteristics are only available for these students.
Source: Student questionnaires. Own calculations.

Furthermore, information given in the parents' questionnaire is used in order to learn more about the students' socio-economic background. However, since only two thirds of the parents completed the questionnaires, it may be that the impression given by the parents' information is not representative for all students' parents.

Table 4C.3 includes information related to the parents' socio-economic background. Most of the questionnaires (two-thirds) were filled in by the students' mothers. In addition, girls' parents more often participate in the survey (61 %) than boys' parents, while the participation rate of fathers is higher for girls (40 %) than for boys (21 %).

Since the sample of responding parents is a selected group one can not directly compare the given information to the children's general information. For example, compared to the children's statements there are fewer persons with an immigrant background in this group.

Table 4C.3: Parental information on socio-economic background

	All parents			Girls' parents			Boys' parents		
	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)
Respondent of questionnaire…									
…female parent	49	0.65	(0.48)	30	0.60	(0.50)	19	0.74	(0.45)
…male parent	49	0.33	(0.47)	30	0.40	(0.50)	19	0.21	(0.42)
…both parents	49	0.02	(0.14)	30	0.00	(0.00)	19	0.05	(0.23)
Family situation:									
Single parent	49	0.31	(0.47)	30	0.33	(0.48)	19	0.26	(0.45)
Number of children*	49	2.24	(1.11)	30	2.40	(1.25)	19	2.00	(0.82)
Grandparents live nearby	49	0.39	(0.49)	30	0.40	(0.50)	19	0.37	(0.50)
Immigration background:									
Female parent native born	48	0.44	(0.50)	30	0.43	(0.50)	18	0.44	(0.51)
Male parent native born	48	0.38	(0.49)	30	0.43	(0.50)	18	0.28	(0.46)
Foreign language at home	49	0.53	(0.50)	30	0.50	(0.51)	19	0.58	(0.51)
Mothers' secondary education:									
…no secondary degree	45	0.13	(0.34)	16	0.06	(0.25)	29	0.17	(0.38)
…lower secondary degree	45	0.51	(0.51)	16	0.56	(0.51)	29	0.48	(0.51)
…intermediate degree	45	0.24	(0.43)	16	0.31	(0.48)	29	0.21	(0.41)
…higher secondary	45	0.11	(0.32)	16	0.06	(0.25)	29	0.14	(0.35)
Fathers' secondary education:									
…no secondary degree	40	0.08	(0.27)	16	0.13	(0.34)	24	0.04	(0.20)
…lower secondary degree	40	0.53	(0.51)	16	0.44	(0.51)	24	0.58	(0.50)
…intermediate degree	40	0.30	(0.46)	16	0.44	(0.51)	24	0.21	(0.41)
…higher secondary	40	0.10	(0.30)	16	0.00	(0.00)	24	0.17	(0.38)
Tertiary degree dummies (reference: no tertiary degree):									
Mother: vocational degree	47	0.57	(0.50)	30	0.53	(0.51)	17	0.65	(0.49)
Father: vocational degree	42	0.69	(0.47)	25	0.68	(0.48)	17	0.71	(0.47)
Mother: higher tertiary	46	0.07	(0.25)	29	0.07	(0.26)	17	0.06	(0.24)
Father: higher tertiary	43	0.05	(0.21)	26	0.08	(0.27)	17	0.00	(0.00)
Mothers' employment status:									
…not employed	47	0.34	(0.48)	29	0.41	(0.50)	18	0.22	(0.43)
…minor employment	47	0.15	(0.36)	29	0.14	(0.35)	18	0.17	(0.38)
…part-time	47	0.23	(0.43)	29	0.14	(0.35)	18	0.39	(0.50)
…full-time	47	0.28	(0.45)	29	0.31	(0.47)	18	0.22	(0.43)
Fathers' employment status:									
…not employed	40	0.15	(0.36)	25	0.16	(0.37)	15	0.13	(0.35)
…minor employment	40	0.03	(0.16)	25	0.04	(0.20)	15	0.00	(0.00)
…part-time	40	0.10	(0.30)	25	0.12	(0.33)	15	0.07	(0.26)
…full-time	40	0.73	(0.45)	25	0.68	(0.48)	15	0.80	(0.41)
*Computer use at work dummies (reference: person does not use computer)***									
Computer use of mother	46	0.24	(0.43)	28	0.18	(0.39)	18	0.33	(0.49)
Computer use of father	44	0.34	(0.48)	26	0.27	(0.45)	18	0.44	(0.51)
*Household income (classified information)****									
Household income***	42	4.76	(2.36)	28	4.82	(2.60)	14	4.64	(1.86)

Note: The calculations only include parents of students observed throughout the year since most background characteristics are only available for these students. *Number of children living in the household. **Among the parents using computers for work 37.5 % indicate to use it for standard software or the internet and 62.5% use special application software. ***Gross household income in categories. If I interpolate between categories, the mean of 4.76 corresponds to EUR 1,730.
Source: Parents' questionnaires. Own calculations.

However, the information confirms the impression of intergenerational immobility: Most of the lower secondary students' responding parents have obtained a lower secondary degree as well. While 57 % of the responding mothers and 69 % of the fathers hold a vocational degree (mostly on the apprenticeship level), there are only few exceptions in which a parent holds a higher tertiary degree (three out of 46 mothers and two out of 43 fathers). The relatively low education of parents goes hand-in-hand with low full-time employment rates (only 28 % of mothers and 73 % of fathers in the available sample are employed) and a low median household income (EUR 1,730)[109].

Table 4C.4: Children's computer use and performance

	All parents			Girls' parents			Boys' parents		
	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)
Do you think it is important that your child attends computer lessons in fifth grade?*	48	1.21	0.46	30	1.30	0.53	18	1.06	0.24
Did you use a computer for the preparation of the child's homework at home (dummy)?	49	0.55	0.54	30	0.50	0.57	19	0.63	0.50
How many hours a day does your child spend on a computer during leisure time?	40	1.19	0.94	24	0.89	0.72	16	1.63	1.08
Did your child's elementary school teacher use a computer to teach (dummy variable)?	49	0.47	0.50	30	0.43	0.50	19	0.53	0.51
What grade did your child have in math last year?*	48	3.42	0.74	30	3.57	0.68	18	3.17	0.79
What grade did your child have in German last year?*	48	3.49	0.72	30	3.52	0.72	18	3.44	0.73
Which overall grade describes your child's performance in elementary school?*	74	2.70	0.77	46	2.76	0.57	28	2.61	1.03

Note: The calculations only include parents of students observed throughout the year since most background characteristics are only available for these students. * Grades refer to the German grading system, where the best grade is 1 (= excellent) and the worst grade is 6 (= failed).
Source: Parents' questionnaires. Own calculations.

Table 4C.4 shows additional parental information on the children's skills and computer use. Nearly half of the responding parents (48 %) support the view that it is important to teach elementary computer skills to fifth graders. However, it might be that the survey respondence rate is higher for those who also support computer studies and this number might thus be upwardly biased. The same is true for the number of parents using a computer to help the student prepare his homework (55 %). Additionally, the average respondent's child spend about one hour of her daily leisure time on the computer. Further information from the parents' questionnaires relates to the computer use at elementary school and elementary school grades. Nearly half of the students' have experienced some computer based elementary school lessons. The overall elementary school performance of the students compared is rather low, which is consistent with the fact that these students attend the lower-level secondary schools.

[109] Household income is an interpolated value from reported income categories (cf. questionnaires). The low value suggests that parents mostly include labour income only.

Appendix 4D: Appendix of Tables of Group Characteristics

Table 4D.1: Contingency tables

Birth country / Group	Co-educational	Mono-educational	Total
Foreign born	12	5	17
German-born	39	18	57
Total	51	23	74

Parents' birth countries / Group	Co-educational	Mono-educational	Total
At least one parent foreign born	35	17	52
Both parents German-born	16	6	22
Total	51	23	74

Language at home / Group	Co-educational	Mono-educational	Total
Foreign language	34	14	48
Only German	17	9	26
Total	51	23	74

Immigrant background / Group	Co-educational	Mono-educational	Total
No immigrant background	12	4	16
Some immigrant background	39	19	58
Total	51	23	74

Grandparent(s) / Group	Co-educational	Mono-educational	Total
Do(es) not live nearby	28	12	40
Live(s) nearby	23	11	34
Total	51	23	74

Parents' education / Group	Co-educational	Mono-educational	Total
Both less than Realschule	22	10	32
At least one parent Realschule or Abitur	21	9	30
Total	43	19	62

Parents' interest / Group	Co-educational	Mono-educational	Total
Show no interest in school	17	8	25
Show interest in school	34	15	49
Total	51	23	74

Source: Student and parents questionnaires. Own calculations.

Table 4D.2: Means of observed variables by group

Groups	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Variables	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
Age of student (in January 2005)	11.42	11.50	11.53	11.64	11.55	11.50
	(0.43)	(0.60)	(0.71)	(0.61)	(0.74)	(0.57)
Immigrant background (Reference: no such immigrant background):						
Student: Born in Germany	0.77	0.58	0.75	0.93	0.81	0.71
	(0.44)	(0.51)	(0.45)	(0.27)	(0.40)	(0.49)
Both parents born in Germany	0.31	0.33	0.08	0.50	0.31	0.14
	(0.48)	(0.49)	(0.29)	(0.52)	(0.48)	(0.38)
Only German language used at home	0.31	0.33	0.17	0.50	0.38	0.43
	(0.48)	(0.49)	(0.39)	(0.52)	(0.50)	(0.53)
Some immigrant background	0.69	0.83	0.92	0.64	0.81	0.86
	(0.48)	(0.39)	(0.29)	(0.50)	(0.40)	(0.38)
Number of siblings …						
…absolute	1.54	1.42	2.08	1.43	1.38	1.43
	(1.33)	(1.00)	(0.90)	(1.28)	(0.81)	(0.98)
…more than 3 years younger	0.54	0.58	0.83	0.50	0.44	0.43
	(0.88)	(0.90)	(1.11)	(0.76)	(0.63)	(0.79)
… 3 or less than 3 years younger	0.23	0.17	0.42	0.50	0.31	0.43
	(0.60)	(0.39)	(0.51)	(0.76)	(0.48)	(0.53)
… older, but at most 3 years	0.31	0.08	0.25	0.21	0.19	0.00
	(0.48)	(0.29)	(0.45)	(0.43)	(0.40)	(0.00)
… more than 3 years older	0.31	0.50	0.33	0.29	0.38	0.43
	(0.48)	(0.80)	(0.65)	(0.47)	(0.72)	(0.53)
Parental education (Reference: higher secondary degree of mother / father):						
	0.23	0.50	0.00	0.07	0.25	0.00
Mother's education: missing value	(0.44)	(0.52)	(0.00)	(0.27)	(0.45)	(0.00)
	0.08	0.08	0.17	0.14	0.19	0.00
Mother's education: no secondary degree	(0.28)	(0.29)	(0.39)	(0.36)	(0.40)	(0.00)
	0.23	0.25	0.50	0.50	0.38	0.57
Mother's education: lower secondary degree	(0.44)	(0.45)	(0.52)	(0.52)	(0.50)	(0.53)
	0.31	0.08	0.17	0.07	0.13	0.43
Mother's education: intermediate degree	(0.48)	(0.29)	(0.39)	(0.27)	(0.34)	(0.53)
	0.23	0.58	0.17	0.07	0.44	0.00
Father's education: missing value	(0.44)	(0.51)	(0.39)	(0.27)	(0.51)	(0.00)
	0.00	0.00	0.00	0.14	0.13	0.00
Father's education: no secondary degree	(0.00)	(0.00)	(0.00)	(0.36)	(0.34)	(0.00)
	0.23	0.25	0.50	0.50	0.19	0.57
Father's education: lower secondary degree	(0.44)	(0.45)	(0.52)	(0.52)	(0.40)	(0.53)
	0.38	0.17	0.17	0.07	0.13	0.43
Father's education: intermediate degree	(0.51)	(0.39)	(0.39)	(0.27)	(0.34)	(0.53)
Family Care:						
Grandparents live nearby (dummy)	0.38	0.42	0.50	0.50	0.63	0.14
	(0.51)	(0.51)	(0.52)	(0.52)	(0.50)	(0.38)
Parental interest	0.85	0.58	0.83	0.43	0.56	0.86
(participation in project dummy)	(0.38)	(0.51)	(0.39)	(0.51)	(0.51)	(0.38)
Performance (in other subjects):*						
Child's overall performance (self reported)	2.92	2.50	2.75	2.71	2.69	2.57
	(0.76)	(0.90)	(0.45)	(0.83)	(0.60)	(1.27)
Child's math performance (self reported)	2.69	2.75	2.67	2.64	3.19	2.71
	(0.85)	(0.97)	(0.65)	(0.93)	(0.83)	(1.25)
Child's German performance (self reported)	3.08	2.67	3.42	3.07	2.50	3.14
	(0.64)	(0.78)	(0.79)	(1.14)	(0.89)	(0.90)
Number of observations	13	12	12	14	16	7

Note: The calculations only include students observed throughout the year. Grey-shaded lines refer to the groups taught by the same teachers. * Self reported performance scales range from 1 (= much better than average) to 5 (much worse than average).
Source: Student and parents questionnaires. Own calculations.

Table 4D.3: Means of observed variables by group for females

Groups	Group 1	Group 2	Group 3	Group 4	Group 5
Variables	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
Age of student (in January 2005)	11.35	11.64	11.38	11.20	11.55
	(0.34)	(0.65)	(0.48)	(0.46)	(0.74)
Immigrant background (Reference: no such immigrant background):					
Student: Born in Germany	0.88	0.33	0.67	0.86	0.81
	(0.35)	(0.52)	(0.50)	(0.38)	(0.40)
Both parents born in Germany	0.50	0.33	0.11	0.43	0.31
	(0.53)	(0.52)	(0.33)	(0.53)	(0.48)
Only German language used at home	0.50	0.33	0.11	0.57	0.38
	(0.53)	(0.52)	(0.33)	(0.53)	(0.50)
Some immigrant background	0.50	0.83	0.89	0.71	0.81
	(0.53)	(0.41)	(0.33)	(0.49)	(0.40)
Number of siblings …					
…absolute	1.75	1.67	2.22	1.43	1.38
	(1.67)	(0.82)	(0.97)	(1.27)	(0.81)
…more than three years younger	0.50	0.50	1.00	0.43	0.44
	(1.07)	(0.55)	(1.22)	(0.53)	(0.63)
… younger, but less than three years	0.25	0.17	0.44	0.57	0.31
	(0.71)	(0.41)	(0.53)	(0.79)	(0.48)
… older, but less than three years	0.38	0.00	0.22	0.29	0.19
	(0.52)	(0.00)	(0.44)	(0.49)	(0.40)
… more than three years older	0.38	0.83	0.33	0.43	0.38
	(0.52)	(0.98)	(0.71)	(0.53)	(0.72)
Parental education (Reference: higher secondary degree of mother / father):					
Mother's education: missing value	0.13	0.33	0.00	0.00	0.25
	(0.35)	(0.52)	(0.00)	(0.00)	(0.45)
Mother's education: no secondary degree	0.13	0.17	0.22	0.14	0.19
	(0.35)	(0.41)	(0.44)	(0.38)	(0.40)
Mother's education: lower secondary degree	0.25	0.33	0.56	0.43	0.38
	(0.46)	(0.52)	(0.53)	(0.53)	(0.50)
Mother's education: intermediate degree	0.38	0.00	0.11	0.00	0.13
	(0.52)	(0.00)	(0.33)	(0.00)	(0.34)
Father's education: missing value	0.25	0.50	0.11	0.14	0.44
	(0.46)	(0.55)	(0.33)	(0.38)	(0.51)
Father's education: no secondary degree	0.00	0.00	0.00	0.00	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.34)
Father's education: lower secondary degree	0.25	0.33	0.56	0.29	0.19
	(0.46)	(0.52)	(0.53)	(0.49)	(0.40)
Father's education: intermediate degree	0.25	0.17	0.11	0.14	0.13
	(0.46)	(0.41)	(0.33)	(0.38)	(0.34)
Family Care:					
Grandparents live nearby (dummy)	0.88	0.50	0.89	0.43	0.56
	(0.35)	(0.55)	(0.33)	(0.53)	(0.51)
Parental interest (participation in project dummy)	0.50	0.33	0.33	0.43	0.63
	(0.53)	(0.52)	(0.50)	(0.53)	(0.50)
*Performance (in other subjects):**					
Child's overall performance (self reported)	3.00	3.00	2.78	2.43	2.69
	(0.53)	(0.63)	(0.44)	(0.53)	(0.60)
Child's math performance (self reported)	3.00	3.33	2.78	2.29	3.19
	(0.53)	(0.82)	(0.67)	(0.76)	(0.83)
Child's German performance (self reported)	3.13	3.00	3.11	3.14	2.50
	(0.64)	(0.63)	(0.60)	(1.35)	(0.89)
Number of observations	8	6	9	7	16

Note: These calculations only include students observed throughout the year. Grey-shaded lines refer to the groups taught by the same teachers. * Self reported performance scales range from 1 (= much better than average) to 5 (much worse than average).
Source: Student and parents questionnaires. Own calculations.

Table 4D.4: Performance in general subjects by group type

Group Type		Mixed-gender groups						Single-gender groups					
Sample	All	All		Girls		Boys		All		Girls		Boys	
Variable	K-W. $\chi^2_{(1)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Comparative performance	0.09	51	2.73 (0.75)	30	2.80 (0.55)	21	2.62 (0.97)	23	2.65 (0.83)	16	2.69 (0.60)	7	2.57 (1.27)
Com. math performance	1.73	51	2.69 (0.84)	30	2.83 (0.75)	21	2.48 (0.93)	23	3.04 (0.98)	16	3.19 (0.83)	7	2.71 (1.25)
Com. German performance	1.65	51	3.06 (0.88)	30	3.10 (0.80)	21	3.00 (1.00)	23	2.70 (0.93)	16	2.50 (0.89)	7	3.14 (0.90)
Expected average grade	0.80	45	2.85 (0.73)	25	2.77 (0.68)	20	2.95 (0.79)	22	2.61 (0.91)	15	2.56 (1.01)	7	2.71 (0.70)
Performance measured at end of term:													
Comparative performance	1.18	51	2.88 (0.52)	30	2.87 (0.57)	21	2.90 (0.44)	22	2.73 (0.77)	15	2.93 (0.70)	7	2.29 (0.76)
Com. math performance	0.15	51	2.89 (0.73)	30	2.97 (0.61)	21	2.79 (0.87)	23	2.74 (0.92)	16	3.00 (0.82)	7	2.14 (0.90)
Expected math grade	3.17*	50	3.16 (0.83)	30	3.25 (0.74)	20	3.01 (0.94)	23	2.83 (1.03)	16	3.07 (1.09)	7	2.29 (0.64)
Com. German performance	4.57**	51	3.21 (0.78)	30	3.23 (0.68)	21	3.17 (0.91)	23	2.78 (0.60)	16	2.75 (0.58)	7	2.86 (0.69)
Expected German grade	4.80**	51	3.42 (0.75)	30	3.35 (0.71)	21	3.51 (0.82)	23	2.98 (0.78)	16	2.88 (0.67)	7	3.21 (0.99)
Change com. performance	0.36	51	0.16 (0.67)	30	0.07 (0.52)	21	0.29 (0.85)	22	0.05 (0.79)	15	0.20 (0.77)	7	-0.29 (0.76)
Performance change:+													
Change math performance	1.76	51	0.21 (0.99)	30	0.13 (0.82)	21	0.31 (1.21)	23	-0.30 (1.15)	16	-0.19 (1.17)	7	-0.57 (1.13)
Change German performance	0.00	51	0.15 (0.98)	30	0.13 (0.68)	21	0.17 (1.32)	23	0.09 (0.79)	16	0.25 (0.68)	7	-0.29 (0.95)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in single-gender and mixed groups. ** Significant at the five percent level. * Significant at the 10 percent level. + The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Appendix 4E: Appendix to the Literature Review for Chapter 4

Table 4E.1: Previous research on single-gender computer or science education in Germany

Study	Major Research Question	Data and Method	Main Conclusions	Main Problem
		Type 1: correlation studies*		
Funken *et al.* (1996)	Is coeducation less favourable for girls' **computer**-related performance (computer use and interest)?	Written questionnaires: 1,128 *Gymnasium* students (11th graders) in the school year 1989/90 in Nordrhein-Westfalen (55 % female; 19 % in all-girls schools, no evidence on boys in single gender schools), analysed by descriptive statistics (mainly Chi-squared tests).	More favourable performance-related outcomes for girls from single gender schools (compared to girls from mixed gender schools); similar outcomes of girls in single gender schools and boys in mixed schools.	Selection to single gender schools.
Meyer (1996)	Do students prefer to be taught in single gender **computer** classes?	Project in the city of *Bremen* in the mid-1990s: 14 boys and 15 girls taught in two single gender computer science classes for half a year in ninth grade of the higher secondary school. Assessment by oral interviews. Qualitative presentation of results.	Boys generally prefer coeducation; mixed statements among girls.	Small size of study. Effects are not assessed in a systematic way (*e.g.* using standardized questionnaires).
Nyssen (1996)	Which factors improve girls' confidence, (professional) interest and performance in selected subjects (with a focus on practical **science classes**)? What difference does single gender education make?	Comprehensive school classes in the state of Nordrhein-Westphalen. Longitudinal observation (questionnaires, interviews, video assessment) of 109 students (50 girls, 59 boys) in six groups (2 all-girls groups taught by female teacher, 2 all-boys groups and 2 coeducational groups taught by male teacher, observed in 5th grade (1991/92) and in 7th grade (1993/94). Qualitative presentation of results.	Results are mainly interpreted in favour for single gender education (especially as concerns girls' performance).	Different treatment measures (gender composition of groups, different teachers and instruction methods) are used simultaneously. Especially, instruction of different groups by different teachers reduces comparability of group outcomes. Selection to groups.

Volmerg *et al.* (1996)	Which factors improve girls' interest and participation in **computer** science? What is the role of coeducation?	Observation (oral interviews) of five computer classes in two high schools (*Gymnasium*) in Bremen during 1989/90-1991/92: 3 female single-gender classes (first year: 13 girls in 11th grade followed until 13th grade; second year: 11 girls in 11th grade, observed until 12th grade, third year 15 girls in 11th grade, observed until 12th grade) and 2 mixed classes (10 and 14 students). Qualitative presentation of results.	Mixed findings; introduction of all-girls computer classes increased (initial) participation of girls in this subject	Comparability of different groups is not assured (taught by different teachers; different contents of lessons). Girls anticipate *expected* outcomes of the intervention study and adjust behaviour.
		Type 2: non–experimental evaluation studies*		
Rost and Pruisken (2000)	What is the impact of single gender education on psychological and social outcomes (including students' self-assessment in **mathematics** and biology) when similar students in single and mixed gender classes are compared?	Questionnaire-based assessment of performance in Catholic private higher secondary schools in 1997: 649 fifth and sixth graders in three single and mixed gender schools (161 girls in an all-girls school, 243 girls in single gender classes in two mixed gender schools, 154 girls in mixed gender classes in the two mixed gender schools, 91 boys in mixed gender classes in the two mixed gender schools). Evaluated by comparison of mean outcomes.	No significant impact of the regime.	Effect is only identified for the very specific (selective) group of Catholic private school students. Selection into different school types (single *vs.* mixed gender).
Häußler and Hoffmann (2002, 1998, 1990) and Hoffmann *et al.* (1997)	Which factors (curricular changes, teacher behaviour, class size and single gender *vs.* coeducation) improve girls' interest, self-concept and achievement in **science** (physics) classes?	Sample of 150 girls and 139 boys in 12 experimental classes from six schools and taught by six different teachers, 103 girls and 64 boys in 7 control classes from two schools and taught by six different teachers. All students are seventh graders in the higher level secondary school (*Gymnasium*) in the state of Schleswig-Holstein in 1992/93. Assessment by several standardized written tests and questionnaires. (Regression adjusted) group differences (and differences-in-differences) are calculated.	Positive impact of single gender education especially on girls' outcomes.	Treatment and control groups are located at different schools. No extensive information on selection into different groups. Regression analysis in Hoffmann *et al.* (1997) may suffer from endogeneity of some of the regressors (*e.g.* self-concept).

		Type 3: natural experiments *		
Baumert (1992)	How does girls' performance and interest in **mathematics** (and other subjects) differ between coexisting mixed and single gender education schools?	Based on a representative survey of German seventh graders in 1968/1969 including standardized German, English and mathematics outcomes: Girls and boys in mixed and single gender schools (each of the four groups includes 2,100-2,900 observations). Compares seventh graders in single gender and mixed higher secondary schools in the time of a regime switch in the organizational school types through analysis of variance.	Evidence for selection into the differently organized school types in regions where both types coexist. Both genders perform significantly better in mathematics in single gender schools than in coeducational schools. Girls' interest in mathematics decreases notably in seventh grade in mixed gender classes.	Natural experiment is not really used to evaluate the causal effect of coeducation (no difference-in-difference approach; no before-after-comparisons).
		Type 4: social experiments *		
Kessels (2002)	Are girls more motivated and self-confident if they are taught in single gender **science** (physics) classes?	Core Sample: 270 compehensive school students (eighth graders) from four schools in *Berlin* (87 girls and 62 boys in coeducational classes; 56 girls and 65 boys in single gender classes). Random assignment into single gender and mixed groups. Each participating teacher instructs at least one single gender and one coeducational class. Assessment of sociodemographic variables and outcomes by standardized questionnaires, evaluated by analysis of variance.	Positive impact of single gender education on girls' motivation and self-confidence.	No information is given on background characteristics of the different groups. There seems to be significant sample attrition.

Note: * The four categorized types of studies are: 1) correlation studies which do not solve the evaluation problem, 2) non-experimental evaluation studies controlling for background characteristics, 3) natural experiments implying exogenous policy changes and 4) social experiments where students are randomly assigned to treatment and control groups

LEBENSLAUF

ANDREA MARIA MÜHLENWEG GEB. WEBER

PERSÖNLICHE DATEN

geboren am 27.03.1977 in Karlsruhe, verheiratet, deutsche Staatsbürgerin

BÜRO

Zentrum für Europäische Wirtschaftsforschung
FB Arbeitsmärkte, Personalmanagement und Soziale Sicherung
L 7,1, 68161 Mannheim
Postfach 10 34 43, 68034 Mannheim

Telefon: 0621 1235 280
Email: muehlenweg@zew.de

BERUFSERFAHRUNG

12 / 2002 – 08 / 2004	Wissenschaftliche Mitarbeiterin des Forschungsbereichs Arbeitsmärkte, Personalmanagement und Soziale Sicherung des Zentrums für Europäische Wirtschaftsforschung (ZEW) Mannheim
06 / 2004 – 09 / 2007	Wissenschaftliche Mitarbeiterin des Lehrstuhls Empirische Wirtschaftsforschung und Angewandte Ökonometrie der Technischen Universität Darmstadt
10 / 2006 – 09 / 2007	Wissenschaftliche Mitarbeiterin des Instituts für Arbeitsökonomik an der Leibniz Universität Hannover
Seit 10 / 2007	Wissenschaftliche Mitarbeiterin des Forschungsbereichs Arbeitsmärkte, Personalmanagement und Soziale Sicherung des Zentrums für Europäische Wirtschaftsforschung (ZEW) Mannheim

BILDUNGSWEG

06 / 1996	Abitur in Durmersheim / Baden-Württemberg
10/1996 – 10/2002	Studium der Volkswirtschaftlehre an der Universität Mannheim
08/1999 – 12/2000	Studium im Studiengang Applied Economics an der University of North Carolina at Greensboro (UNCG), USA
09/1998	Vordiplom an der Universität Mannheim
12 / 2000	Master of Arts in Applied Economics (M.A.), University of North Carolina at Greensboro (UNCG), USA
10 / 2002	Diplom-Volkswirtin, Universität Mannheim, Schwerpunkte: Ökonometrie, Finanzwissenschaft, Arbeitsmarktökonomik
06 / 2003 – 11 / 2007	Doktorandin der Volkswirtschaftslehre unter der Betreuung von Prof. Dr. Patrick Puhani (Erstgutachter) und Prof. Dr. Horst Entorf (Zweitgutachter)

TÄTIGKEITEN WÄHREND DES STUDIUMS

04 / 2002 – 11 / 2002 und 03 / 1998 - 10 / 1998	Praktika und Tätigkeit als Ungeprüfte Wissenschaftliche Hilfskraft in der Abteilung Arbeitsmärkte, Personalmanagement und Soziale Sicherung des ZEW Mannheim
09/1998 – 10 / 1998	Praktikum in der Wertpapierabteilung der Landeszentralbank Stuttgart

MITARBEIT AN DRITTMITTELFINANZIERTEN FORSCHUNGSPROJEKTEN

10 / 2006 – heute	The Economics and Politics of Employment, Migration and Social Justice, gefördert durch die Anglo-German Foundation (Leitung des Teilprojektes: Patrick Puhani, Leibniz Universität Hannover; Koordinator: Christian Dustmann, University College London)
12 / 2004 – 01 / 2005	Technische Assistenz im Rahmen der Evaluation aktiver Arbeitsmarktprogramme in Serbien, Beschäftigungsprojekt der Weltbank und des Britischen Instituts für internationale Entwicklung (Leitung technische Assistenz: Markus Frölich, Michael Lechner, Patrick Puhani, Universität St. Gallen)
09 / 2003 – 8 / 2004	Die Entwicklung und Verteilung des Vermögens privater Haushalte unter besonderer Berücksichtigung des Produktivvermögens, für das Bundesministerium für Gesundheit und Soziale Sicherung (Projektleiter: Peter Westerheide, ZEW)
04 / 2002 – 09 / 2003	Qualifikation und Arbeitsmarkterfolg in Deutschland und Frankreich: Der Einfluss von bildungs- und familienpolitischen Maßnahmen im Vergleich, gefördert durch die Fritz Thyssen Stiftung (Projektleiterin: Miriam Beblo, ZEW)

LEHRERFAHRUNG AN DER TECHNISCHEN UNIVERSITÄT DARMSTADT

Sommer 2007	Betreuung des Seminars zur empirischen Wirtschaftsforschung
WS 2006/2007	Übung zur Ökonometrie (Hauptstudium)
Sommer 2006	Betreuung des Seminars zur empirischen Wirtschaftsforschung
Sommer 2005	Übung zur Statistik (Grundstudium)
WS 2004/2005	Betreuung des Seminars zur empirischen Wirtschaftsforschung

VERÖFFENTLICHUNGEN

Referierte Fachzeitschriften und Buchkapitel:

2007
Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Early Educational Effects of Age of School Entry in Germany, *Empirical Economics* 32, 359-386 (mit Patrick Puhani).

2005
Education and Wage Inequality in Germany: Review of the Empirical Literature, in: Rita Asplund und Erling Barth (Hrg.): *Education and Wage Inequality in Europe, A Literature Review*, ETLA Sarja B 212 Series Helsinki, 167-202 (mit Andreas Ammermüller).

2004
Ältere Arbeitnehmer, in: Tobias Hagen und Alexander Spermann (Hrg.): *Auswirkungen der Hartz-Gesetze – Methodische Ansätze zu einer Evaluierung*, ZEW Wirtschaftsanalysen, 227-240.

Diskussionspapiere:

2007
Persistence of the School Entry Age Effect in a System of Flexible Tracking, IZA Discussion Paper, Nr. 2965, Bonn (mit Patrick Puhani).

An Evaluation of Single and Mixed Gender Computer Science Classes, Diskussionspapiere des Fachbereichs Wirtschaftswissenschaften, Universität Hannover, Nr. 369, Hannover.

2006
Educational Effects of Alternative Secondary School Tracking Regimes in Germany, Diskussionspapiere des Fachbereichs Wirtschaftswissenschaften, Universität Hannover, Nr. 353, Hannover.

Fängt der frühe Vogel den Wurm? Eine empirische Analyse des kausalen Effekts des Einschulungsalters auf den schulischen Erfolg in Deutschland, Diskussionspapiere des Fachbereichs Wirtschaftswissenschaften, Universität Hannover, Nr. 336, Hannover (mit Patrick Puhani).

2005
Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany, IZA Discussion Paper, Nr. 1827, Bonn (mit Patrick Puhani).

Educational Attainment and Returns to Education in Germany, ZEW Discussion Paper, Nr. 05-17, Mannheim (mit Andreas Ammermüller).

2004
Wann kehren junge Mütter auf den Arbeitsmarkt zurück? Eine Verweildaueranalyse für Deutschland, ZEW Discussion Paper, Nr. 04-08, Mannheim.

2003
Education and Wage Inequality in Germany – A Review of the Empirical Literature, ZEW Discussion Paper, Nr. 3-29, Mannheim (mit Andreas Ammermüller).

Employment of Mothers After Childbirth: A French-German Comparison, ZEW Discussion Paper, Nr. 03-50, Mannheim (mit Charlotte Lauer).

Gutachten:

2005
Die Entwicklung und Verteilung des Vermögens privater Haushalte unter besonderer Berücksichtigung des Produktivvermögens, Abschlussbericht zum Forschungsauftrag des Bundesministeriums für Gesundheit und Soziale Sicherung, Mannheim (mit Andreas Ammermüller und Peter Westerheide).

VORTRÄGE (AUSWAHL)

- Jahrestagung des Vereins für Socialpolitik, Bayreuth, 2006
- ZEW-Rhein-Main-Neckar Arbeitsmarktseminar, Mannheim, 2005
- IZA European Summer School in Labor Economics, Buch/Ammersee, 2005
- 19th meeting of the European Economic Association (EEA), Madrid, 2004
- 18th Annual Conference of the European Society for Population Economics (ESPE), Bergen/Norwegen, 2004
- IZA European Summer School in Labor Economics, Buch/Ammersee, 2004

AUSZEICHNUNG

BAK (Basel-Economics) Preis für Empirische Wirtschaftsforschung 2003

GREMIENTÄTIGKEITEN

Seit Sommer 2006	Stellvertretende Vertreterin der Wissenschaftlichen Mitarbeitenden im Fachbereichsrat des Fachbereichs „Rechts- und Wirtschaftswissenschaften“ an der Technischen Universität Darmstadt
Seit WS 06/07	Mitglied des Direktoriums des Instituts für Volkswirtschaftslehre an der Technischen Universität Darmstadt
Sommersemester 2007	Vertreterin der Wissenschaftlichen Mitarbeitenden in einer Berufungskommission für Empirische Wirtschaftsforschung an der Technischen Universität Darmstadt