

Institutional Effects in the Production of Education:

Citation for published version (APA):

Ammermüller, A. (2007). *Institutional Effects in the Production of Education*: [Doctoral Thesis, Maastricht University]. ROA. <https://doi.org/10.26481/dis.20070530aa>

Document status and date:

Published: 01/01/2007

DOI:

[10.26481/dis.20070530aa](https://doi.org/10.26481/dis.20070530aa)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**INSTITUTIONAL EFFECTS IN THE
PRODUCTION OF EDUCATION:**

Evidence from
European Schooling Systems

© Andreas Konrad Hermann Ammermüller, Maastricht 2007

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing, from author.

Published by ROA
Postbus 616
6200 MD Maastricht

ISBN: 978-90-5321-455-8

Printed in the Netherlands by Océ Business Services

INSTITUTIONAL EFFECTS IN THE PRODUCTION OF EDUCATION:

Evidence from
European Schooling Systems

DISSERTATION

to obtain the degree of Doctor at
the Maastricht University,
on the authority of the Rector Magnificus,
Prof.dr. G.P.M.F. Mols
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday 30 May 2007, at 12 hours

by

Andreas Konrad Hermann Ammermüller

Supervisor:
Prof. dr. J.A.M. Heijke

Co-supervisor:
Dr. S. Schim van der Loeff

Assessment Committee:
Prof. dr. W. Groot (voorzitter)
Prof. dr. L. Borghans
Prof. dr. L. Woessmann (LMU München, Germany)

Acknowledgments

The interest in the topic of education economics arose during an internship in the course of my studies in Maastricht with Ludger Woessmann in Kiel. It was followed by a master thesis on educational production under the supervision of Hans Heijke. Besides writing a joint paper, I have stayed in regular contact with both and profited from various comments and discussions on different aspects of my work.

The work on this thesis began when I joined the Centre for European Economic Research (ZEW) at the end of 2002. My first project „Education and Wage Inequality in Europe” (EDWIN) allowed me to pursue my research interests in the field of the economics of education. I owe thanks to Charlotte Lauer, who was the project leader at ZEW, and to the project partners, who invited us to all sorts of fascinating places in Europe and contributed to the discussion of the research. Charlotte provided me with valuable feedback to the empirical analysis of this study and also introduced me to scientific working.

The outline of the thesis was developed together with my supervisor Hans Heijke and draws mostly on work that was completed during the EDWIN project. Rewriting the papers and adding the theoretical and methodological part in combination with a long process of commenting took most of the time. I would like to thank Hans Heijke for his great effort and sincere interest in the supervision of the thesis. I would also like to thank my second supervisor Sybrand Schim van der Loeff for his scrutiny in reading and commenting in particular the empirical chapters. Moreover, I owe thanks to Jozien Hendrikx who gave her kind support in administrative matters.

My colleagues contributed to a friendly working environment at ZEW and were always the first persons I could bother with any kind of problem. I would like to thank my parents for their lasting support. Finally, I am thankful to Britta, who constitutes my better half since my studies and supported me not only by reading my work but in particular by her devotion and love.

Andreas Ammermüller, spring 2007

Contents

Acknowledgments	v
1 Introduction	1
1.1 Background and rationale	2
1.1.1 The role of education for economic and social outcomes	2
1.1.2 Evidence on the effect of school quality on educational outcomes	4
1.1.3 Research questions	12
1.2 Methodological approach	13
1.3 Main findings	14
1.4 Plan of the thesis	15
2 Educational production	17
2.1 Theoretical aspects of educational production	17
2.1.1 Theory of educational production	17
2.1.2 The role of institutions in the schooling systems	24
2.2 Empirical aspects of educational production	28
2.2.1 Measures of education	28
2.2.2 Data limitations and estimation problems	34
3 Explaining the difference in student performance between countries	41
3.1 Introduction	41
3.2 The dataset	43
3.2.1 Distribution of test scores	45
3.3 Determinants of reading proficiency scores	48
3.3.1 The educational production function	48
3.3.2 The estimated effects	50
3.4 Explaining the test score gap	52
3.4.1 Oaxaca-Blinder decomposition	53
3.4.2 Juhn-Murphy-Pierce decomposition	56
3.5 Conclusion	63
4 Educational opportunities and the role of schooling institutions	65
4.1 Introduction	65
4.2 Data	68
4.2.1 Imputation of missing values	68
4.2.2 Comparability of the studies	69
4.3 The effect of student background on student performance	71

4.3.1	Estimation outline	71
4.3.2	Estimation results	73
4.4	The role of institutions	78
4.4.1	Schooling systems	78
4.4.2	Theoretical aspects	79
4.4.3	Estimation strategy	80
4.4.4	Empirical evidence	85
4.4.5	Robustness tests	91
4.5	Conclusion	93
5 Synthesis and conclusion		95
5.1	Summary of findings	96
5.2	Synthesis	98
5.3	Policy conclusions	100
5.4	Research prospects	101
Literature		103
Appendix (chapter 3)		117
Appendix (chapter 4)		127
Summary in Dutch		149
Curriculum Vitae		153
Roa Dissertation Series		155

1

Introduction

In today's knowledge-based economy, in which the accumulation of new information grows rapidly and yesterday's know-how becomes quickly outdated, the basic knowledge and skills acquired in schools are growing ever more important. Later education builds on knowledge obtained earlier, or as Heckman puts it, "[e]arly learning begets later learning" (Heckman, 2000, p.5). The specific content taught in higher education instead may turn into obsolete wisdom rapidly and help little in acquiring new knowledge. Therefore, the amount and quality of schooling is of utmost importance for the accumulation of a society's human capital and serves as a foundation on which further education can prosper. Furthermore, student performance can best be influenced and inequalities in the provision of high quality education be corrected while students are still young (Heckman, 2005).

Education was already recognized as an important economic factor by the classic economists Adam Smith and Alfred Marshall. In his 1776 masterpiece *Wealth of Nations*, Smith formulated that "[a] man educated at the expence of much labour and time . . . may be compared to [an] expensive machin[e]" (p. 118). Marshall acknowledged the importance of capital invested in workers in his work *Principles of Economics* (1890), but denied the existence of human capital as a form of an economy's wealth. Education was hence not incorporated explicitly into economic theory before the human capital theory was stated by Schultz in 1961 and further works by Becker (1964) and Mincer (1974), which manifested the branch of *economics of education*. In modern endogenous growth theories (Romer, 1990; Mankiw et al., 1992), the significance of human capital for economic growth has been set at comparable levels to physical capital and labor.

The importance attributed to education has risen in recent years since human capital accumulated through education is regarded to be the essential production factor in the modern knowledge-based economy. Knowledge that results from the investment in human capital is ever more important for designing new products and services and thus for enhancing economic growth and welfare (Drucker, 1969). Due to the positive externalities of education and its public good nature, the state is still the major provider of compulsory education, as it prefers to exert influence over this significant determinant of contemporary society. In political campaigns, education has become a central issue and remains one of the top points on the political agenda since the publication of the comparable test results from the

Programme of International Student Assessment (PISA) (OECD, 2001; 2004). The below average performance of economic frontrunners like Germany and Switzerland has been considered as a national fiasco in those countries and has led to controversial discussions on how to improve school quality. As the industrialized countries stay in ever-closer competition to each other through the effects of globalization and liberalization, they fear to lose their competitive edge. Therefore, "European countries are becoming more and more aware that they must devote unremitting attention to the level of knowledge and skills of their populations if they wish to maintain their global competitiveness" (Heijke and Muysken, 2000, p. xvii). The quality of the educational system should hence be a major point on the political agenda (cf. Borghans and Heijke, 2005; Sachverständigenrat, 2004).

This thesis is concerned with the production of education at the level of compulsory schooling, in particular with the performance of students in lower secondary education. Its main focus lies on the theoretical and empirical analysis of the link between student background and school quality including schooling institutions and student achievement. Besides analyzing the relationship between the input factors of the schooling system and student performance, the study seeks to explain differences in the degree of equality of educational opportunities by schooling institutions.

1.1 Background and rationale

1.1.1 *The role of education for economic and social outcomes*

Education is well-known to be a major determinant of wages, income and non-market benefits on the individual level and of economic growth and the income distribution on an aggregate level. The importance that is directed at school education in particular is supported by a growing literature that stresses the extent to which schooling quality affects the earnings prospects of students and economic growth. This section underlines the importance of education for economic and social outcomes and addresses the role of both student achievement and the degree of equality of educational opportunity.

The benefits of education

The positive link between education and wages and hence income is one of the best documented relationships in the economic literature (cf. Card, 1999).¹ The size of the causal effect of education on wages is, however, subject to an ongoing controversy. Student achievement measured by tests in literacy or numeracy is as well strongly linked to wages (e.g. McIntosh and Vignoles, 2001; Murnane et al., 1995; 2000). Besides wages, education determines also non-market outcomes such as health (e.g. Berger and Leigh, 1989) and life expectancy (e.g.

1. See also Asplund and Barth (2005) for a review of the empirical literature for Europe and Ammermueller and Weber (2005) for recent evidence for Germany.

Feldman et al., 1989).² As is also shown in this study, parents' education has a strong impact on the educational achievement of children and hence on their income but also on their health (e.g. Glewwe, 1999). On an aggregate level, education affects economic growth and is a major input factor in endogenous growth models (e.g. Romer, 1990; Mankiw et al., 1992; Aghion and Howitt, 1998). The quality of education can alternatively be used as an explanatory factor (Barro, 2001; Hanushek and Kimko, 2000), e.g. for weighting the years of schooling in growth regressions (cf. Gundlach and Woessmann, 2004). The income distribution is as well closely linked to the distributions of education and student achievement (Tinbergen, 1975; Blau and Kahn, 1996; 2005). Besides economic outcomes, education has been shown to further social values such as democracy, political stability and a low level of criminality (cf. McMahon, 1999; Lochner and Moretti, 2004). A detailed literature review on the evidence on the link between school quality and student achievement is given in the following section while section 2.2.1 discusses the link between education and economic outcomes in relationship to the measurement of student achievement.

The role of educational opportunities

One important aspect of education that lies at the focus of this study is the degree of equality of educational opportunity, i.e. in how far the background of students determines their educational performance.³ The intergenerational mobility of human capital and hence of income depends on the degree of equality in educational opportunities (e.g. Björklund and Jäntti, 1997; Dearden et al., 1997). Accordingly, social mobility within societies is largely determined by educational opportunities. A higher degree of equality of educational opportunity can avoid costly redistribution through tax and benefit systems later on (Nickell, 2004). From an economic point of view, e.g. an extremely low degree of equality in opportunities would imply that investment in human capital depends less on innate ability and more on social origin, which leads to a non-optimal investment in human capital of individuals. Thus, educational opportunities are likely to affect also significant economic outcomes like economic growth. Mejia and St-Pierre (2005) show in a theoretical model that a low degree of equality of educational opportunity leads to a lower average level of human capital accumulation and higher wage inequality. Schooling systems play a central role in shaping the determinants of student performance with regard to the background of students. A schooling system that is characterized by a strong use of streaming, i.e. a system that groups students by ability, might be more prone to producing a heavy dependence of performance on social background and hence lead to lower intergenerational mobility. Evidence on the Finnish comprehensive school reform supports this hypothesis (Pekkarinen et al., 2006). The role of schooling institutions in determining educational opportunities is at the heart of this thesis and is analyzed theoretically in Chapter 2 and empirically in Chapter 4.

2. See Haveman and Wolfe (2000) and Vila (2000) for overview articles on non-market benefits of education.

3. The degree of equality of educational opportunity is defined as the impact of student background on student achievement in this study. A similar concept is used to measure the degree of intergenerational mobility of income or education (cf. Blanden et al., 2005).

1.1.2 Evidence on the effect of school quality on educational outcomes

This section discusses the empirical evidence on the link between school quality measures and further school characteristics such as resources, teacher characteristics, school types and institutions and educational outcomes of students. This state of the art literature review shows where the focus of research is on and summarizes the available evidence. The insights gained from the literature review are then used to identify the research questions of interest which are addressed in this thesis.⁴

Resources

The main resource measures in this field are direct expenditure per student and teachers per student. Their effects on educational outcomes have been estimated in a series of studies dating back to the famous Coleman Report from 1966. From a theoretical point of view, students should clearly benefit from more resources in the form of better equipment and smaller classes. In view of the production function literature, more resources imply an outward shift of the production possibilities frontier, which should always lead to a higher production outcome. Especially low performing students, who need special attention, should benefit when optimal class size differs with student ability (Lazear, 2001). However, if teachers overestimate the benefits of their time allocated to individual students relative to the class and tend to spend less time teaching the entire class when they have a small class, then a reduction of class size might actually be detrimental to student performance (Borghans, 2002).

Despite the clear-cut theoretical implications, the evidence on resource effects is ambiguous and actually suggests a low effectiveness of resources in improving student performance. This view was mainly determined by several influential reviews and meta-analysis by Hanushek (1986a, 1986b, 1989), although the applied method of combining the results from various different studies in the meta-analyses is questionable (cf. Hedges et al., 1994). A main conclusion is that resources alone are no guarantee for a higher quality of education. Other inputs into the educational process such as the incentive framework are essential as well (Hanushek, 2003). The conclusion that resources have at best a limited impact on educational outcomes remains hotly debated, though (cf. Hanushek, 2003; Krueger, 2003).

Recent empirical evidence also suggests that an increase in resources, as measured by expenditure per student, is mostly ineffective in improving student achievement in empirical studies or to have very small positive effects in relation to costs (see Betts (2001) for evidence on the U.S., Dolton et al. (2004) for a summary of the European literature, Levacic and Vignoles (2002) for England and Woessmann (2003) for further international evidence). This is also the case in school-level analyses for England (Gibbons, 2002) and Finland (Häkkinen et al., 2003), where a reduction in spending during the 1990s has been evaluated in the latter case. Studies using country-level data (e.g. Lee and Barro, 1997) or the variation across countries (e.g. Fuchs and Woessmann, 2006) instead often find a positive impact of educational

4. See Woessmann (2006a) for a recent review on education policies in Europe and Woessmann (2002) for a general discussion of student performance and human capital.

expenditure on student performance. It is however doubtful in how far these studies are able to control for country-specific effects like the degree of development and thus run the risk of misinterpreting their findings. Dolton and Vignoles (1999) show that expenditure per pupil by the Local Education Authorities in England is negatively linked to the probability of being unemployed when leaving school. More specific studies that examine the use of computers at schools find no positive effect on student achievement (Angrist and Lavy, 2002; Fuchs and Woessmann, 2004). The evidence on spending programs targeted at specific schools and students is mixed. For England, Machin et al. (2004) and Machin and McNally (2004) find positive short run effects of the “Excellence in Cities” program on student achievement and positive long run effects of the “National Literacy Project”. For two subsidy programs targeted at disadvantaged students in the Netherlands, no positive effects on student attainment are found (Leuven et al., 2004). The specific design and target group of spending programs seem thus to determine their success.

The other central resource indicator examined in the empirical literature is the amount of teachers measured by student-teacher ratio or class size. Here again, the evidence points to no or weak effects of resources, especially in studies using student or school level data. While class size is a more direct measure of resources that students receive, it is more likely to be endogenous due to selection processes within schools (cf. Dolton and Vignoles, 2000). Analyses of international student performance studies indicate the existence of strong selection effects both between and within schools (Woessmann and West, 2006). Even the use of school-fixed effects model and instrumental variables (e.g. Ammermueller et al., 2005) or of sequential models using panel data (e.g. Dearden et al., 2002) cannot prove beneficial effects of smaller classes for students. Examples of no or a low impact are Dearden et al. (2002) for England and Wales, Woessmann and West (2006) for most OECD countries and Ammermueller et al. (2005) for Eastern Europe. For the U.S., Hoxby (2000) uses population variation to identify class size effects in primary schools but even rules out modest effects of class size. Using longitudinal data, Betts (2001) cannot establish any link between school resources and educational attainment but finds positive impacts of resources on black women’s wages but none for white women. Bressoux et al. (2005) instead find significant negative effects of class size on third grade entry test scores in France. Figlio (1999) estimates significant but tiny effects of student-teacher ratios using a less restrictive educational production function. For Israeli public schools, Angrist and Lavy (1999) find substantial negative class size effects for fourth and fifth graders but not for third graders. For private independent schools in the UK, a negative relationship between the student-teacher ratio and average examination results at the school level can be observed (Graddy and Stevens, 2003).

Studies using country-level data often find a negative link between student-teacher ratios and test scores and dropout rates (e.g. Lee and Barro, 1997). The probability of continuing education increases only by very little when students are in smaller classes in England and Wales (Dustmann et al., 2002). Brunello and Checchi (2005) find that lower student-teacher ratios are positively correlated with higher educational attainment and are a substitute for lower parental education in Italian regions. Even when studies find positive effects of smaller classes for students, the estimated effects are rather small and are confronted with extremely high

costs of reducing class size. It has to be noted, however, that these results are only valid for the ranges of class size and student-teacher ratios that are observed. It is quite likely that there are upper and lower thresholds above which the impact may well be strong.

Thus, the evidence on the impact of resources on educational performance is rather puzzling. No clear pattern emerges from the empirical evidence on observational data. For the impact of expenditure per student, studies using country-level data are more likely to find positive effects than micro-level analysis, but the effects remain small. In general, more significant effects of smaller classes and other resource measures are found for students in lower grades and for disadvantaged students. This does not necessarily mean that resources do not matter for the average student. Both the empirical evidence and theoretical considerations underline the great difficulty in consistently estimating the causal relationship between class size, student-teacher ratio or financial resources and student performance. It remains questionable which method to use to solve the problem of separating the quality of the school from the quality of students as well as the role of prior inputs, and the data sets generally entail limited information (see Todd and Wolpin, 2003, Hanushek, 2003). This calls for social experiments. The Tennessee STAR (Student Teacher Achievement Ratio) experiment randomly allocated children in kindergarten and grades 1 to 3 to small and large classes. Children placed in small classes performed better in standardized tests than children in large classes (Krueger, 1999). These achievement gains persisted throughout the schooling career and could even be traced to the results of college entrance examinations (Krueger and Whitmore, 1999). However, it is open to debate whether the effects justify a reduction in class size (cf. Hanushek, 2003; Krueger, 2003). Moreover, social experiments like project STAR suffer from specific drawbacks and their conclusions remain doubtful. Individuals might act differently when they take part in a social experiment. The observed results cannot be generalized then. Moreover, it has to be noted again that the results are only valid for the ranges of class size and student-teacher ratios and resource levels that are observed. Nevertheless, the main conclusion based on the evidence available so far is that resources alone are no guarantee for a higher quality of education.

Teacher characteristics

Not only is the number of teachers important, but also and above all the quality of their teaching, i.e. their ability to raise pupils' achievement. Again, the first difficulty consists in measuring teachers' or teaching quality. Typically, the empirical literature distinguishes three indicators to proxy teachers' quality: the formal qualification of teachers, their experience and also their earnings, the latter being an indicator of the incentive system and of the ability of the school system to attract good teachers. The question which arises is in how far these observable characteristics are able to capture teaching quality. The second difficulty consists in separating teachers' contribution to student achievement from mere selection effects, i.e. student placement or teacher assignment to specific classes depending on ability and achievement.

A higher quantity and quality of education and training of teachers can be expected to improve teaching. Experience should help teachers to transmit the content more effectively to students. However, the age of teachers increases along with experience and may restrict a close relationship between teachers and students as well as lead to a lower motivation of teachers.

The empirical evidence suggests a small but significant effect of the education of teachers on student performance in several micro-level studies. Especially high achieving students seem to benefit from well-trained teachers (Ammermueller, 2006; Bressoux et al., 2005; Goldhaber and Brewer, 1997). Other studies instead find no evidence of an impact of higher formal education of teachers, especially for the U.S. (Rivkin et al., 2005; Hanushek and Rivkin, 2003). Teachers' sex has no effect on student performance on average but there is some evidence for positive student-teacher gender interaction effects (Ammermueller and Dolton, 2006; Dee, 2005).

The experience and hence age of teachers seems to have a small positive but declining effect on student performance (e.g. Hanushek, 2003). Thus, there appear to be significant gains in the first years of experience and smaller gains over the following years in the career of teachers. This might be the combined results of two distinct phenomena. First, young teachers might go through a learning-on-the-job phase at the beginning of their career and improve their teaching skills in the first years of their career. This implies that allocating novice teachers to schools or classes where a high proportion of pupils face learning difficulties, as happens in France due to the teachers' assignment system, is expected to be counter-productive. Second, a fraction of young teachers might discover that they are not suited for teaching and leave the profession in their first working years. Rivkin et al. (2005) present evidence for the U.S. that the dominant effect is the learning-on-the-job effect and not selective exits from the profession. Overall, the relationship between teachers' education and experience is not strong and some studies even find no significant effects of teachers' characteristics at all (e.g. Haegeland et al. (2005) for Norway).

Teachers' salaries are a composite of pay for certain attributes, typically experience, education and other attributes such as marital status or school location, cost-of-living etc., but are generally not related to student performance. Studies examining the relationship between teachers' salary level and student achievement find no strong evidence that salaries are a good measure of teacher quality. The link appears to be very loose, especially if the determining attributes (experience and education) are controlled for. However, evidence from several social experiments in Israel suggests that performance-related incentive pay for teachers does raise educational performance of students and is relatively cost-effective compared to cash bonuses (Lavy, 2003; 2004). Using tests as a measure to raise teacher quality in public schools in some U.S. states does lead to higher teacher wages but not to an improvement in teacher quality (Angrist and Guryan, 2005).

Moreover, the level of teachers' wages compared to wages in other professions might determine the attractiveness of the profession and the ability of the school system to attract good teachers. Recent trends show that the recruitment of high quality teachers is becoming more

difficult in most countries (Dolton, 2004; Hanushek and Rivkin, 2003), partly due to a shift of teachers' wages compared to other professions. Some studies, though, argue that working conditions and non-pecuniary factors are more important than salaries for attracting good teachers (Hanushek and Rivkin, 2003). Goldhaber et al. (1999) find that teachers' behavior is more important than actual resource inputs.

Overall, the relationship between teachers' characteristics and student achievement is rather loose. A new study by Rivkin et al. (2005) shows, using a panel data set, that there are indeed large differences in teachers' quality, i.e. in the ability of teachers to raise student achievement, but that only little of the variation in teacher quality is explained by observable characteristics such as education or experience. Thus, the loose relationship emerging from the empirical literature might simply result from the fact that the available indicators are poor proxies for teachers' quality.

Type of school and class composition

School types differ in several ways. In most countries, a certain fraction of students attends private schools instead of public schools. Private schools can either be managed privately, be partly financed privately or both. Moreover, school types can differ by their orientation. In countries that use streaming in higher secondary education, schools prepare students either for further vocational training or tertiary education. The optimal timing of streaming in the schooling system may thereby depend on both the benefits of more specialized teaching and the costs of an early selection of students (Brunello et al., 2004). Besides the selection of students into different school types, students may select themselves or be allocated into schools and classes depending on their characteristics. This affects the composition of classes and thus the direct learning environment of students. Both the selection of students and the exposure to the same environment such as having the same teacher give rise to the reflection problem (Manski, 1993), which makes it difficult to identify and disentangle peer effects. School policies may range from creating very homogenous to heterogeneous classes with respect to the characteristics of students. The rationale for creating homogenous peer groups is that teachers can address the class as a whole instead of having to consider the specific situation of each student. Advocates of heterogeneous peer groups instead claim that students learn from each other and weak students are not left behind. Therefore, the variation in the class composition as well as the average performance level is an important issue when analyzing peer effects (Glewwe, 1997). Besides the distribution of performance, student's sex and ethnic origin are also part of class composition and could affect student achievement (Hoxby, 2003).

The empirical literature focuses on various aspects of differences between schools, depending on schooling policy and special traits of the schooling system. In the U.S., two major points of discussion are independent religious schools and school vouchers. In many European countries, there is great discussion on the grouping of students but little empirical evidence so far.

Catholic schools in the U.S. seem to have positive effects on high school and college graduation rates and future wages mainly for urban minorities, while there are only modest effects for urban whites and none for suburban students (Neal, 1997). Catholic schools seem thus only to make a difference when the alternative public schools are of a low quality. Concerning non-market behavior, the attendance of religious private schools reduces some behavior such as teen sexual activity and arrests but has no effect on other negative behavior such as the use of soft drugs (Figlio and Ludwig, 2000). A large program of school vouchers for private schools in a U.S. state was found to benefit students in math but not in reading (Rouse, 1998). The design and the target group of private school voucher programs are decisive for its equity effects, as shown in general equilibrium simulations (Nechyba, 2000). In Colombia, a school voucher program that covered half the costs of private education proved to be successful in increasing student performance in the short run (Angrist et al., 2002) and secondary school completion rates in the long run (Angrist et al., 2004).

When studies control for a large amount of student and school characteristics, no difference is found between private and public schools for most European countries (Vandenberghe and Robin, 2004). One exception is Belgium, where private schools seem to benefit student performance. However, Belgium is a special case in Europe since it is the only country apart from the Netherlands where more than half of all students attend private schools (OECD, 2002). Brunello and Rocco (2004) discuss instead the possibility that private schools in Italy provide a lower quality of education than public schools and can find some empirical evidence in favor of their hypothesis. When distinguishing between operating and funding, Woessmann (2006b) finds in an international cross-sectional comparison that public funding is correlated with higher while public operation with lower average student performance.

The difference in student performance between school types that differ in their orientation is huge in countries like Germany that heavily apply streaming (Baumert et al., 2001). However, as students are selected into school types according to their ability, it is difficult to measure the pure school type effect. Longitudinal studies for Britain suggest that students placed in higher streams benefit from streaming while students in lower streams do not (Robertson and Symons, 1996). The most able students seem to benefit from a more selective school system, while there is no evidence on significant negative effects for low ability students (Galindo-Rueda and Vignoles, 2004).⁵ The average performance of students appears to increase when selection within the school system is reduced. This is the main result of an evaluation of a change in the admission rules in Northern Ireland at the end of the 1980's (Maurin and McNally, 2006). Figlio and Page (2000) report no negative effects of streaming for low ability children in the U.S. as well, while Slavin (1990) concludes from a review of studies that ability grouping is ineffective in general. In Italy, graduating from a general high school is associated with a higher probability of attending university while technical schools improve the quality of the school-to-work transition of graduates (Cappellari, 2004). The access to private schools and a selective school system is likely to affect educational opportunities within regions or countries as well. Hanushek and Woessmann (2006) establish a

5. Pischke and Manning (2006) show that selection leads to biased estimates in these studies analyzing the British comprehensive school reform, though.

positive link between streaming and a greater variation of test scores, using variation both between countries and over student age. A Swedish school reform that increased compulsory schooling and abolished early selection is investigated by Meghir and Palme (2005). They find positive effects of the reform on educational attainment and earnings for students from a lower social background but not for students whose parents are highly educated. A similar comprehensive school reform in Finland can be linked with an increase in intergenerational income mobility (Pekkarinen et al., 2006).

The composition of classes is found to be important regarding the average performance level of peers but not the heterogeneity of student performance in Texan school classes (Hanushek et al, 2003). Rangvid (2003) confirms these findings for Denmark even when additional data sources are used to alleviate a potential omitted variable bias. Similar results emerge from a study on peer effects in Austrian secondary schools (Schneeweis and Winter-Ebmer, 2005). Moreover, these studies show that peer level effects are strongest for low performing students. The evidence suggests that a policy of mixing abilities would help to maximize average student performance. Hoxby (2003) finds only little evidence on non-linear peer effects but can show that peer effects are not only present for achievement of students but as well for race and gender. Ammermueller and Pischke (2006) provide evidence on significant average peer effects in primary schools for several European countries. An evaluation of a U.S. busing program (METCO), in which mostly black students are transferred to suburban schools, shows only very slight negative effects of the new class composition on the performance of local students (Angrist and Lang, 2002). Using data from the social experiment STAR, Graham (2004) disentangles peer effects by using the difference in excess variance between small and large classes. Thereby, he can identify the strength of the social multiplier, which indicates the overall gain from a higher achievement of students through positive peer effects and feedbacks. The estimated size of the social multiplier ranges between 1.07 and 2.31 for math and 1.05 and 3.07 for reading, which would imply great gains for students from improved achievement of peers.

School types and class composition seem to affect both the average and the distribution of student performance. The evidence on the effects is fairly weak and points to significant effects of private schools only for students from a disadvantaged background and to positive effects of selection only for the most able students. The average performance of peers has a strong positive influence on one's own performance in most studies, while there is little evidence of any effect of class heterogeneity.

Institutional setting

The institutional setting of a schooling system sets the framework in which actors involved in the process of educational production act. It mainly determines the freedom of choice and of taking decisions and creates incentives for students, teachers, parents and other interest groups (Bishop and Woessmann, 2004). The main institutional features discussed in the literature are central (exit) exams, school and teacher autonomy and remuneration systems. School type and class composition also belong to the institutional setting but have been

discussed separately. Compared to investments in resources, changes in institutions are less cost intensive. However, they may only be effective in the long run.

A consistent finding from the literature is that the presence of central exit exams tends to improve student performance. This is both shown in studies using the variation between countries (e.g. Bishop, 1997; 1999; Woessmann, 2005) and between regions within one country (Jürges et al., 2005). External standards also positively affect the attitude of teachers and their effort in teaching students (see Jürges et al., 2004 for Germany). The evidence on the link between external standards and educational opportunities is mixed. While some standard tests help to close the gap between students from different social backgrounds, others have no effect (Bishop and Mane, 2005).

School autonomy is found to be beneficial for student performance, especially when external standards are set (e.g. Fuchs and Woessmann, 2006; Robin and Sprietsma, 2003). The age at which students enter school is negatively correlated with later educational attainment in Germany but has no significant impact on attainment after controlling for selection bias (Fertig and Kluge, 2005). Puhani and Weber (2005) instead find small positive effects of later school entry age on student performance using an instrumental variable approach. For Sweden, small positive effects of a later school starting age on attainment are found as well (Fredriksson and Öckert, 2004). Besides deciding on when students start school, students can be enrolled earlier and be taught before the official start of school. Distinguishing between students shows that an earlier enrolment yields small positive effects for disadvantaged students but none for other students in the Netherlands (Leuven et al., 2006). The placement in kindergarten in Germany yields similar results, namely that immigrant students benefit while there is no effect for other students (Spiess et al., 2003).

The evidence on the effect of the institutional setting on student achievement is rather promising compared to the evidence on resources and teacher characteristics. Setting the right incentives, especially in form of central exams, seems to improve student performance given resources and seems to be a prerequisite for an effective use of further spending on education.

Overall, the literature finds that student characteristics and family background are the most important determinants of student achievement, while the evidence on school quality effects is mixed and rather weak.⁶ One pattern that emerges from several studies is that some measures of school quality benefit disadvantaged students while having no effect for the average student. This underlines the importance of distinguishing between students with different characteristics. However, very few studies focus on how differences in schooling outcomes between students, which vary greatly between countries (e.g. Woessmann, 2004), are affected by institutional features.

6. Student background factors are not discussed explicitly in the review of the literature but their strong impact is the finding of all studies that control for differences in student background. The recent literature in this area focuses on the causality of impact factors such as parents' education (cf. Black et al., 2005), which is not directly related to the issues investigated here.

1.1.3 Research questions

One of the few common results from the literature presented above is the fact that student background is the most important determinant of student performance and has a far greater impact than school quality measures and other school characteristics. Moreover, there is strong evidence of a great variation in the degree of equality of educational opportunity between countries, i.e. in the effects of student background on performance. The evidence suggests that an increase in educational resources through higher spending does not necessarily improve the quality of education. Especially in relation to its costs, measures relying on the use of more resources are not effective in improving the quality of education and hence student performance. Institutional frameworks, which drive the incentives of actors involved in the production of education – students, teachers, parents and administrators –, are a more promising tool to improve both efficiency and equity of education. This holds all the more since changes in institutions, although they may only be effective in the longer run, are less cost intensive than other measures. Moreover, the focus of research has been on the effects of resources and shifts only recently towards the role of institutions in the process of educational production. With few exceptions, the economic literature is concerned with the average level of student achievement as the primary outcome of education. The evidence on the equality of opportunity, i.e. the effect of student background, instead has been mostly descriptive so far. Some of the most recent literature investigates the determinants of educational opportunities and income mobility but focuses only on one characteristic of a schooling system or on a specific reform.⁷

Therefore, this thesis focuses on institutional effects in educational production. It investigates how several institutional features like the differentiation of the school system, instruction time and school autonomy are related to the equality of educational opportunity. Explaining differences in educational opportunity by schooling institutions is the main contribution of this study to the literature. The approach followed is both theoretical and empirical, although the emphasis lies on the empirical analysis. The study addresses the following questions: what determines student performance and what is the contribution of various input factors to student achievement as the outcome of the production of education in different countries? What is the degree of the different dimensions of equality of educational opportunity in selected countries? What is the role of institutions for educational opportunities? How can a sound institutional framework of schooling systems contribute to increases in both efficiency and equity in educational production?

The analysis conducted in the line of this thesis strives for adding new insights to the production of education, especially concerning the role of institutions and the degree of educational opportunity in schooling systems. Its aim is to provide theoretical and empirical evidence on the functioning of educational systems in order to foster student performance and provide equal opportunities in the provision of high quality education.

7. An exception is Schuetz et al. (2005) who consider several institutional features but only use cross-sectional variation to identify the effects.

1.2 *Methodological approach*

In order to answer the research questions, both data from the Programme of International Student Assessment (PISA) and Progress in International Reading Literacy (PIRLS) studies for several countries are analyzed and the variation of institutions both within and between countries is exploited.

Before the empirical analysis, the study provides the reader with the methodological framework for the later parts of the study. First, a theoretical model of educational production is developed which takes account of the heterogeneity of student background. Thereby, the interaction between schooling institutions and student background effects can be analyzed in the model. Second, the empirical aspects of educational production are discussed. The study presents several measures of educational attainment and compares their usefulness for economic analyses. Furthermore, it addresses the limitations of student performance data and difficulties in the estimation of production functions and identifies possible solutions to overcome these problems. It follows an in-depth empirical analysis of institutional effects.

The first empirical analysis investigates the determinants of student performance in lower secondary education in PISA 2000 for two different schooling systems. The countries of choice are Finland and Germany. While Finland is one of the leading countries in recent performance tests and features a low degree of inequality in the distribution of test scores, German students perform below OECD average in PISA and their test scores are widely distributed. Moreover, Germany relies on streaming in its secondary school system, whereas there is only one school type in Finland. The effects of student background, the use of resources and institutional features that vary within the countries are estimated by educational production functions using enriched student-level data. The differences in the level and spread of test scores are then explained by differences in the distribution of characteristics and endowments in the two schooling systems and the differential effect they have on student performance. A cross-sectional version of the Juhn-Murphy-Pierce decomposition method is used to determine the contribution of different characteristics and the estimated effects to explaining the difference in the test score distributions between the two countries. The study analyzes institutions that vary within the countries and addresses the role of streaming. This first part of the empirical analysis hence identifies the significance of different inputs for explaining differences in the distribution of test scores between countries but only exploits the variation of inputs within countries. The comparison is only performed between two countries to go into greater detail regarding the differences in student performance and to decompose the test score gap.

The second empirical investigation continues the analysis from the former part and scrutinizes the role of institutions for educational opportunities. It exploits the variation of institutions between several countries to explain the large differences in student background effects between countries. The study describes the degree of equality of educational opportunity for several countries at two important stages of a student's schooling career, the end of primary education (grade four) and the end of lower secondary education (grade nine/ten). The effects of student background on student performance are estimated and compared both between

countries and over student age using PISA 2000 and PIRLS 2001 student-level data. The possible interactions between the different dimensions of educational opportunities, e.g. the effects of social origin and parental attitude, and institutions are explored. The empirical analysis tests the hypotheses on the determinants of educational opportunities from the theoretical model. As institutions develop over a long time span within countries, they may be strongly linked to a country's culture and tradition. Relating educational opportunities simply to current institutions in a cross-section of countries would yield biased estimators since we cannot control for country-specific effects. Therefore, the empirical analysis exploits changes in educational opportunities and institutions over student age. Thereby, country effects can be largely controlled for and a consistent relationship between educational opportunities and institutions can be estimated. A synthesis of the findings and policy recommendations conclude the thesis.

1.3 *Main findings*

The main findings refer in particular to the empirical analyses, which build on the review of the literature and the theoretical model of educational production and educational opportunities. The comparison of Finnish and German students' test scores shows that Finnish students perform much better than German students at the age of fifteen in PISA, especially at the lower end of the test score distribution. While the gap between the two countries amounts to 30 points for the highest performing decile of students, it reaches over 75 test score points for the lowest decile. Test scores are determined to a higher degree by observable characteristics in Germany than in Finland. Especially the impact of social background is higher in the streamed schooling system of Germany. Differences in student background and resources can hardly explain any part of the difference in performance. Instead, German students are endowed with a higher social background than Finnish students on average. Only for the lowest deciles of the student distribution, student background is more favorable for Finnish than for German students and can explain a part of the score gap. However, the use of resources and the transformation of student characteristics into student performance are estimated to be more effective in Finland than in Germany and can hence explain a part of the test score gap. The main difference between the educational processes in the two countries is the degree of equality of educational opportunity, which is much higher in Finland than in Germany.

The main hypotheses developed from the theoretical model are supported by the empirical findings. Schooling institutions appear to be linked to the degree of equality of educational opportunities in schooling systems across countries. A high differentiation of the schooling system, which is either due to streaming or a large private school sector, decreases educational opportunities for students with a lower social background. The stronger the use of streaming, as indicated by the number of school types, the greater is the effect. A larger share of students attending private schools steadily increases the impact of social background on student performance. The amount of instruction time instead positively affects the degree of equality of educational opportunity. A higher degree of school autonomy benefits students

whose parents have a high commitment to their child's education more relative to students whose parents care less about educational achievement.

1.4 *Plan of the thesis*

The various aspects of educational production are discussed in Chapter 2. First an introduction to the theory of educational production and the role of institutions is given. A theoretical model is developed which allows for an interaction between schooling institutions and educational opportunities of students. Then a deeper insight into empirical aspects is provided. This includes a comparison of different measures of educational attainment and a discussion of data limitations and estimation problems that arise in empirical educational production functions. The third chapter constitutes the first part of the empirical analysis and investigates educational production and the contribution of various input factors to schooling outcomes. The analysis is conducted for the countries of Finland and Germany and hence contrasts educational production in two very distinct schooling systems. The fourth chapter continues the line of reasoning by investigating the role of institutions for educational opportunities using the variation across countries and over student age. Finally, Chapter 5 summarizes the results and provides a synthesis of the insights gained in the former chapters. The chapter closes with conclusions and policy recommendations from the findings.

2

Educational production

The greatest part of educational production takes place during compulsory schooling (cf. Neal and Johnson, 1996). While higher secondary and tertiary education deepen the general knowledge and enable students to specialize in certain subjects, the basic cognitive skills are acquired in primary and lower secondary education or even before. They constitute the foundation of human capital which is essential for the further accumulation of skills (cf. Heckman, 2000). This study hence focuses on the process of educational production during compulsory schooling. The parts of this chapter, which serves as a methodological introduction to the following analytical chapters, discuss both theoretical and empirical aspects of educational production in schooling. First, a theoretical model is developed and the interaction between schooling institutions and educational opportunity is analyzed by means of the model. Second, the most prominent measures of education are evaluated and the shortcomings of student performance data and problems arising in the estimation of educational production functions are discussed.

2.1 Theoretical aspects of educational production

Although the emphasis of this study is on the empirical analysis, the theoretical aspects of educational production are discussed in the following parts to provide a sound basis for the further empirical chapters.¹ A theoretical model on the production of cognitive achievement in schools follows a short introduction into the theory of educational production. The model is developed to study the effects of institutional features in schooling systems on the degree of equality of educational opportunity.

2.1.1 *Theory of educational production*

A thorough comparison of student performance in the schooling systems of different countries presupposes the knowledge of the process by which education is produced. Educational production functions provide a means of understanding the production process by estimating the effects that various inputs have on student achievement. Since the groundbreaking works

1. I would like to thank Nicole Gürtzgen and Marisa Hidalgo-Hidalgo for helpful comments on the theoretical model.

by Schultz (1961) and Becker (1964), human capital did not only enter utility and earnings functions as an important input factor but its production was also explained by theoretical models of educational production. Two of the first basic models on the production of human capital are found in Becker (1967) and Ben-Porath (1967), which link the life-cycle of earnings to the investment in human capital. The analysis of educational production builds upon the production function literature, which models the production of various goods by input factors such as labor and capital. As production functions seek to explain the production of a homogeneous good, most of the economic studies on this subject restrain their analysis to a single outcome of education, in particular cognitive skills of students. Other positive effects of education upon students, such as the social values conveyed to students, are consciously neglected. This study as well considers only cognitive skills as measured by student performance tests.²

Due to the complex nature of educational production, theoretical models focus mostly on a selection of input factors. The most prominent factors include parental characteristics and early home environment, which are often classified as “informal” factors along with further factors from outside the school. Research on these factors is especially popular among sociologists and educational researchers. Economists concentrated their work on the effects of resources in producing cognitive skills, which belong to “formal” factors from inside the school. More recently, the role of institutions has been included in theoretical models. Hoxby (1999) examines the effect of the level of school financing and Epple and Romano (1998) and Brunello and Rocco (2004) consider the quality of private and public schools.³ The role and timing of ability tracking is discussed in recent papers by Epple et al. (2002) and Brunello et al. (2004). Bishop and Woessmann (2004) present a more general model that includes the effects of different institutions such as central exams and school autonomy.

However, the theoretical models reviewed above focus on effectiveness aspects in the production of human capital, i.e. on the impact of institutions on average student achievement. Theoretical papers on equity aspects of the impact of institutions are scarce. Epple et al. (2002) consider the impact of tracking on the relative performance of students from different socio-economic status. Hidalgo-Hidalgo (2005) highlights the role of peer effects for human capital accumulation and college attendance from both an efficiency and equity point of view. Besides the average accumulation of human capital, she uses the gap in college attendance between students from rich and poor households as a second criterion. Schuetz et al. (2005) provide a theoretical model in which they study the impact of pre-school enrollment and streaming and conclude that long pre-school education and late tracking are beneficial for equality of opportunity.

Another strain of the literature considers the relationship between economic inequality, human capital accumulation and macroeconomic outcomes (eg. Galor and Zeira, 1993; de Gregorio, 1996; Benabou, 1996, 2000). The aim of these studies is to explain the observed link

2. Test scores developed to measure cognitive skills might reflect the outcome of both cognitive and non-cognitive skills (Borghans et al., 2007; Cunha et al., 2005; Heckman, 2005).

3. School quality is defined over student achievement and is different from school resources.

between the income distribution and per-capita income on the macroeconomic level through the access to education. A further theoretical study suggests that a higher degree of equality of educational opportunity leads to a higher average level of human capital formation and lower wage inequality (Mejia and St-Pierre, 2005). The degree of educational opportunity is determined exogenously to the model, though.

The theoretical model outlined below adds to the theoretical literature by focusing on how various institutions affect differences in performance between students from a different social background in compulsory schooling. However, the model does not seek to explain the effects on the absolute level of student achievement. The equality of educational opportunity is defined as the strength of the impact of student background upon student performance in the context of this study.

The education production function

Educational production depends on a wide range of input factors. A general model of the production of cognitive skills in school looks like the following:

$$(2.1) \quad T = T(E, B, V, R).$$

The level of cognitive skills T of a student depends on student effort E , student and family background B , parental attitude or commitment of parents to their child's education V and school and teacher characteristics R . Student background B combines the input factors of the home environment like resources and family situation except for parents' commitment to education V . The institutions that are expected to have a direct effect on average student achievement such as central exams are not included in the model while the institutions that are more likely to affect educational opportunities and in which we are interested in this context take an effect on the production process. Student ability is not an explicit production factor in the model but can be thought of as an exogenous factor that is independent of student background. As the production of education is a cumulative process which starts in early childhood and continues up to the date when the educational achievement is measured, the inputs are not indexed by time but instead represent the cumulative inputs over time. Equation (2.1) presents a very general model of educational production without specifying the form of the production function, i.e. the way in which the different input factors contribute to skill formation.

To model the production process, a multiplicative specification that allows for the interaction of different input factors on the basis of a Cobb-Douglas production function like in Bishop and Woessmann (2004) is used. The simplified educational production function that is used for the theoretical model can be written as follows:

$$(2.2) \quad T_j = (E)^\alpha \left((1 - N)\phi B_j \right)^\beta (\phi V)^\gamma (NR)^\delta.$$

The index $j = H, L$ distinguishes students from a high social background and from a low social background, with $B_H > B_L$. The other variables in equation (2.2) are not indexed, which implies that students are homogeneous regarding their own ability and that resources and school characteristics R are identical for all students. The variable N indicates the share of time students learn in school and $1-N$ the time students learn at home. The way how student background B_j and parents' commitment V enter the production function are represented by the parameters φ and ϕ , respectively, which depend on the institutional setting. The parameter φ states e.g. the available choice of schooling institutions, while ϕ is determined by the possibilities of parents to influence their child's education. The educational production function depicted by equation (2.2) takes hence not only account of the input factors but considers as well the process by which inputs are transformed into student achievement. It is important to note that the parameter φ describes only the effect of the available choice of schooling, e.g. the possibility to attend private schools, and not the effect of the production in the different institutions themselves. Moreover, the model is set up to study the impact of a single parameter at a time and not the simultaneous effect of several changes in the framework that affect the production of education. The impact of institutional features of schooling systems through the parameters in equation (2.2) is discussed in more detail in the following section. The elasticities of the respective inputs are indicated by α, β, χ and δ . Under the given assumptions the heterogeneity in the model stems only from differences in student background. This is sufficient to study the impact of input factors on the equality of educational opportunity. The model implicitly assumes that student ability, which is not included explicitly in the model, is homogenous conditional on student background. This assumption is made in most of the theoretical literature.

Students and parents are the decisive actors in this model and determine the endogenous variables student effort E and parental commitment V by maximizing their respective utility. Both students and parents act as agents in this model. Since the variable R is not used in the maximization process of the model, we can set $R=1$. The production function then further simplifies to:

$$(2.3) \quad T_j = E^\alpha \left((1-N)\varphi B_j \right)^\beta (\phi V)^\chi N^\delta.$$

The set-up of this production function accounts for both production in school and at home and allows for a heterogeneous student background. Thereby the model of the production function presented by equation (2.3) contributes to the theoretical literature. It is assumed that the function has decreasing returns to scale, i.e. $\alpha + \beta + \chi + \delta < 1$.

What we are actually concerned with, however, is not the absolute level of student achievement T_j but rather the difference in performance between students from a high and a low social background. The equation of interest is the following measure of the degree of equality of educational opportunity:

$$(2.4) \quad \text{EO} = 1 / (T_H - T_L),$$

where *EO* stands for the equality of educational opportunity.⁴ Whatever increases the difference in performance between students from a high and low social background is associated with a decrease in educational opportunities. This definition of educational opportunity relates to the concept of ‘level the playing field’ described by Roemer (1998) for the equality of opportunity. This concept states that all individuals should receive the same preconditions during their period of formation. The following parts lead through this simple model of educational production and describe the maximization process by students and parents and the resulting equilibrium.

Student maximization

Students themselves are the most important actors in the process of education. It is their effort and learning ability, which determines the outcome of educational production strongly. While the ability of students is given, they can choose a certain level of effort. On the one hand, student effort leads to higher student achievement and thereby to a higher self-esteem and respect by teachers, fellow students and parents, to a higher probability to participate in higher education, to a better access to university and finally to higher potential life-time earnings. Moreover, health risks are lower for better educated people. The benefits of students G_s include therefore both pecuniary incentives as well as intrinsic benefits. On the other hand, student effort E is associated with costs C_s , which arise from the time that is invested in learning instead of playing with friends or investing the time in some other activity. Furthermore, students may be harassed by fellow students when investing a high level of effort in learning.

The utility function of students U_s reflects the trade-off between the benefits G_s and costs C_s of student effort. It can be approximated by the following equation:

$$(2.5) \quad U_s = G_s - C_s = \tau T_j - c_s E^\mu, \quad \mu > 1.$$

The parameter τ is a summary measure of the benefits that arise from student achievement T_j . The parameter increases for example when the discount rate of life-time earnings decreases or when educational achievement is more closely related to higher earnings or a better reputation. The cost factor c_s represents the disutility from student effort. The elasticity of costs with respect to effort of students μ is greater than one because the marginal cost of effort is assumed to rise as effort increases. This assumption seems reasonable when considering the limited time available to students and the likely increasing harassment by fellow students.

$$(2.6) \quad E = \left[\frac{\tau \alpha}{c_s \mu} \left((1 - N) \phi B_j \right)^\beta (\phi V)^\alpha N^\delta \right]^{\frac{1}{\mu - \alpha}}.$$

4. The equality of educational opportunity could alternatively be defined as the ratio T_L / T_H or the absolute value of the difference $T_L - T_H$. Since we are not interested in quantitative results, the exact definition of *EO* is not important.

When students maximize their utility, they choose the level of effort depicted in equation (2.6). Since $\mu > \alpha$ by definition, the cost factor c_s and the elasticity μ are negatively related to student effort, while all other factors are positively related to student effort. The effect of the time spent learning in school N depends on the size of the elasticities β and δ . The optimal level of student effort is higher for students from a high social background than from a low social background.

Parent maximization

Parents are also important actors in the process of educational production. In this model of educational opportunity, only parents differ while students and schools are homogeneous. Therefore, it is straightforward to consider the maximization process of parents as well. The following equation states the utility function of parents:

$$(2.7) \quad U_P = G_P - C_P = \psi T_j - c_{Pj} V^\eta, \quad \eta > 1.$$

Parents reap direct benefits from the performance of their children T_j because they are proud of their offspring, have to worry less about their son's or daughter's future or expect to benefit directly from the expected earnings of their children later on. The parameter ψ indicates how strongly student achievement enters the utility function U_P . The costs of parents depend on their commitment V to the education of their children and on a cost factor c_{Pj} . Costs consist of the time, patience and money parents need to spend to improve their children's educational achievement, either through helping their children directly, by talking to teachers or school official or by paying for additional teaching. The cost factor c_{Pj} differs for parents with a high and a low social status. It can be assumed that parents with a high social status, who are themselves highly educated, have a high innate sympathy towards education and more resources so that it is less costly for them to exert a given level of V compared to parents with a lower social status, i.e. $c_{PH} < c_{PL}$. The elasticity η is the elasticity of cost with respect to parents' commitment. As in equation (2.5), it is assumed that the marginal cost of commitment rises as commitment increases.

$$(2.8) \quad V = \left[\frac{\psi \chi}{c_{Pj} \eta} E^\alpha \left((1-N) \phi B_j \right)^\beta \phi^\chi N^\delta \right]^{\frac{1}{\eta-\chi}}$$

The optimal level of parents' commitment V , which maximizes the utility of parents, is shown in equation (2.8). The cost factor and elasticity are negatively related to parents' commitment, while all other variables and parameters affect V positively except for the time spent in school, whose effect is ambiguous. Parents with a high social status have a higher optimal level of V than parents with a low social status, since $\eta > \chi$.

Equilibrium

In equilibrium, both student and parent utility are maximized simultaneously. This assumes that education is a process where both students and parents take account of each other and consider the behavior of others when determining their own level of effort/commitment. By solving equations (2.6) and (2.8) the endogenous variables E and V can be expressed only as a function of the exogenous variables, parameters and elasticities. In equilibrium, the level of student effort is represented by:

$$(2.9) \quad E^* = \left[\left(\frac{\tau\alpha}{c_s\mu} \right)^{\eta-\chi} \left(\frac{\psi\chi}{c_{pj}\eta} \right)^\chi \left((1-N)\phi B_j \right)^{\beta\eta} \phi^{\chi\eta} N^{\delta\eta} \right]^{\frac{1}{\eta\mu-\alpha\eta-\chi\mu}} .$$

Both the cost factors and the elasticities of costs enter the equation negatively, i.e. an increase in any of these factors leads to a lower level of student effort in equilibrium. The level of student effort is higher for students from a high than from a low social background. The same holds for the level of parents' commitment, which is presented in the following equation:

$$(2.10) \quad V^* = \left[\left(\frac{\tau\alpha}{c_s\mu} \right)^\alpha \left(\frac{\psi\chi}{c_{pj}\eta} \right)^{\mu-\alpha} \left((1-N)\phi B_j \right)^{\beta\mu} \phi^{\chi\mu} N^{\delta\mu} \right]^{\frac{1}{\eta\mu-\alpha\eta-\chi\mu}} .$$

Equations (2.9) and (2.10) can then be used to solve for the equilibrium level of student achievement by substituting the equations into equation (2.3). This yields:

$$(2.11) \quad T_j^* = \left[\left(\frac{\tau\alpha}{c_s\mu} \right)^{\alpha\eta} \left(\frac{\psi\chi}{c_{pj}\eta} \right)^{\mu\chi} \left((1-N)\phi B_j \right)^{\beta\eta\mu} \phi^{\chi\eta\mu} N^{\delta\eta\mu} \right]^{\frac{1}{\eta\mu-\alpha\eta-\chi\mu}} .$$

Again, the cost factors and elasticities of cost are negatively related to student achievement, as could be expected. The relationship between time spent in school and student achievement depends on the relative size of the elasticities β and δ . The parameters τ , ψ , ϕ and ϕ all enter equation (2.11) positively. This holds as well for student background B_j . The necessary condition that is needed for the model to yield the qualitative conclusions on the institutional effects later on is $\eta(\mu-\alpha) > \chi\mu$. This condition develops from equation (2.11) and is fulfilled under the sufficient conditions $\eta, \mu > 1$ and $\alpha + \beta + \chi + \delta < 1$.⁵ However, what we are actually interested in is not the level of student achievement but the difference in the equilibrium level of the achievement of students from a high and a low social background. This measure of educational opportunities is shown in equation (2.12).

5. To see this, divide the necessary condition by μ . This yields $\eta(1-\alpha/\mu) > \chi$. The maximum value for χ under the sufficient conditions is marginally smaller than $1-\alpha$. Since the left hand side of the equation is always larger than $1-\alpha$ when η and μ are larger than 1, the necessary condition is fulfilled.

$$(2.12) \quad EO^* = 1 / (T_H^* - T_L^*)$$

Given that $B_H > B_L$ and $c_{PH} < c_{PL}$, students from a high social background have a higher level of student achievement than students from a low social background, i.e. $T_H^* > T_L^*$. The following section now investigates how various institutional features of the schooling system affect the degree of equality of educational opportunity.

2.1.2 *The role of institutions in the schooling systems*

The focus of the analysis is on schooling institutions for which there is little evidence on a link to average student performance. However, the institutional features are suspected of being closely related to the degree of educational opportunities of students in a schooling system. Therefore, the following four institutional features are considered: the amount of instruction time, the size of the private school sector, the use of streaming and the degree of school autonomy. The discussion on educational opportunities focuses on these institutional features in many countries. Examples are the call for the abolition of streaming and a higher instruction time in order to reduce the difference in performance between students from different social backgrounds in Germany and the debates on private schools and school autonomy in the UK and the Netherlands (e.g. Economist, 2005; 2006a; 2006b). Other institutions like central exit exams are neglected instead. For these institutions, the debate focuses rather on the impact on the average level of student performance. Moreover, it has been shown that the impact of central exit exams does not differ substantially for students from a different background (cf. Woessmann, 2005). The theoretical model is designed to analyze the effects of institutions only separately, i.e. under the ceteris paribus assumption. No interaction between institutional features is to be explained by this model.

Streaming

Most studies on streaming, ability grouping or tracking are from sociologists and educational researchers. The focus lies on the effect of streaming on students with different abilities and different social backgrounds. The main argument in favor of ability grouping is that teachers can adapt the pace of instruction and the curriculum to the ability level of their students (Slavin, 1990). Low achievers receive more support and high achievers benefit from the stimulation of other high achievers (Feldhusen, 1989). Arguments that oppose the use of streaming state that students in low tracks fall further behind in relative performance due to a lower quality of instruction for low tracks (e.g. Gamoran, 1989; Oakes, 1985; Schuetz et al., 2005). Moreover, streaming might discriminate against students from a lower social background (e.g. Braddock, 1990; Rosenbaum, 1976). This is the case when the selection of students into tracks or school types also depends on social background, conditional on previous achievement, which seems to be the case in many tracking systems (e.g. Bos et al., 2003; Figlio, 2005). This is possible when either schools belonging to higher school types rather select students from a high social background because they expect them to be better students or students are restricted in their choice because their parents cannot or are unwilling to provide the access to higher school types (see section 2.2.2 for a discussion of selection effects). Theory

and the empirical evidence seem to suggest that the use of streaming has at best no effect on average educational achievement but might well lead to a decrease of the degree of equality of educational opportunities (e.g. Epple et al., 2002; Hanushek and Woessmann, 2006).

In the theoretical model presented above, we can test the effect of streaming on students from a different background but not on students with different ability because we assume that student ability is homogeneous. Since the theoretical arguments on the effects of ability grouping on performance go in both directions, we restrict the effect of streaming to the choice that is available to students in form of different school types. This is also the measure of streaming in the empirical analysis in Chapter 4. Selection between school types is likely to be more important than streaming within schools in this context because the school environment also differs in the former case. In the theoretical framework, the use of streaming widens the choice of institutions which students can attend and would hence lead to an increase in ϕ . Whether certain schools or school types benefit children through peer effects, different resources or other factors remains open. The argument here is that children and parents have a greater choice when there are several school types and that the freedom of choice depends on social background. Moreover, the importance of parents' commitment for their child's achievement increases in schooling systems with streaming, i.e. ϕ increases because parents can influence the type of school that their children attend:

$$(2.13) \quad \frac{\partial \phi}{\partial \text{stream.}} > 0, \quad \frac{\partial \phi}{\partial \text{stream.}} > 0.$$

As ϕ directly interacts with student background B_j and ϕ interacts with parents' commitment V , student achievement increases stronger for students from a high social background than for other students. Hence, the difference between students from a high and a low social background increases and the degree of equality of educational opportunities decreases.

$$(2.14) \quad \frac{\partial \text{EO}^*}{\partial \text{stream.}} < 0.$$

The intuition of this result is that students from a higher social background profit more from a greater choice of institutions because they have better access to institutions. This is due to their already higher achievement at the time of selection and might be furthered by the discrimination of students from a low social background in the selection process. Moreover, parents of students from a higher social background have a higher level of commitment to their child's education in equilibrium, giving greater support to their child for attending better schools.

Instruction time

The time students spend in school and are actually instructed by teachers might limit the effect of other differences between students on their performance. When a student's cognitive achievement is both determined by learning at school and at home, a higher instruction time should decrease the part of educational production that takes place at home (Todd and Wolpin, 2004).

In the theoretical model, an increase in instruction time leads to an increase in N :

$$(2.15) \quad \frac{\partial N}{\partial \text{instr.time}} > 0.$$

The variable N enters equation (2.11) for the equilibrium level of student achievement twice: on the one hand, the increase in N leads to a decrease in the importance of social background B_j . On the other hand, a larger N increases the production of education in school. From equation (2.11) we know that an increase in N would lead to an increase in EO^* under the following condition:

$$(2.16) \quad \beta N > \delta(1 - N).$$

This condition is fulfilled if $N \geq 0.5$ and $\beta > \delta$, which is likely because students spend most of their learning time in school and the importance of student background is found to be greater than the importance of schools in all empirical studies. If condition (2.16) holds, what we can reasonably assume, then an increase in instruction time leads to a greater degree of equality of educational opportunity:

$$(2.17) \quad \frac{\partial EO^*}{\partial \text{instr.time}} > 0.$$

Size of private school sector

A certain fraction of students attends private schools for at least part of their compulsory schooling. These may be either privately financed, independent or both (Vandenberghe and Robin, 2004). The access to private institutions differs between types of institutions and countries. Some private schools levy fees while others are free of charge. Common to the vast majority of private schools is that parents and their children have to apply to schools and these can choose their students. Therefore, the access to private institutions depends on the commitment of parents to their child's education. Moreover, when schools select their students, they might discriminate against students from a lower background because these are expected to receive less support at home and might imply more trouble for teachers and principals. Hence, the share of the private school sector might decrease educational opportunities due to a lower access for disadvantaged students. This argument only holds for the relatively low shares of private schools observed in most European countries, with the exception of Belgium. When the share is sufficiently high, a further increase would imply that also disadvantaged students are given the possibility to attend private schools. However, it is likely that a large private school sector is highly segregated by means of tuition fees.

An increase in the size of the private school sector has the same effect as streaming. It leads to an increase in the parameters φ and ϕ :

$$(2.18) \quad \frac{\partial \varphi}{\partial \text{private}} > 0, \quad \frac{\partial \phi}{\partial \text{private}} > 0.$$

An increase in φ or ϕ implies that the background of students becomes more important and leads to a larger improvement of achievement of students from a high social background than from a low social background. Hence, the degree of equality of educational opportunity decreases:

$$(2.19) \quad \frac{\partial \text{EO}^*}{\partial \text{private}} < 0.$$

School autonomy

In most countries, decision making is shared between schools, local education boards and the regional or central school administration. Decisions are taken on hiring and firing of teachers, salaries, the curriculum, school budgets etc. School autonomy, i.e. the degree of freedom that schools have in taking these decisions, differs strongly between countries. While the relationship between school autonomy and student performance is ambiguous and depends on the type of school autonomy (cf. Ammermueller et al., 2005; Bishop and Woessmann, 2004), the link to educational opportunities should be clear-cut. Therefore, we do not distinguish between the levels where decisions are taken like the state, regions, communities, schools and teacher. When the distribution of responsibility is further decentralized, parents have both a stronger incentive to take part in decision making and a greater possibility of influencing teachers or principals instead of some unknown regional official. Therefore, greater school autonomy should increase the effect of parental attitude on student performance.

An increase in school autonomy and in the possibility of parents to influence their child's school education is reflected by an increase in ϕ .

$$(2.20) \quad \frac{\partial \phi}{\partial \text{autonomy}} > 0.$$

In the theoretical model, a higher ϕ implies that the contribution of parents' commitment V to student achievement increases. The optimal level of V is higher for students from a high social background because they have a lower cost of commitment. Therefore, the equilibrium level of student achievement increases stronger for students from a high social background than for student from a low social background when the degree of autonomy increases.

$$(2.21) \quad \frac{\partial EO^*}{\partial \text{autonomy}} < 0.$$

The degree of educational opportunity decreases because it is easier for parents to affect decision making and take direct influence at schools in a decentralized school system. Parents of students from a high social background are more likely to do so because their equilibrium level of commitment is higher, which benefits their children relatively more.

The theoretical model developed in this section displays a process of educational production that is driven by the behavior of students and parents. The heterogeneity of parents allows for studying the impact of institutional features of the schooling system on the degree of equality of educational opportunity. It can be shown that streaming, a large private school sector, low instruction time and high school autonomy lead to an increase in the importance of student background in determining student achievement and hence to a decrease of the degree of equality of educational opportunity. The theoretical model is limited in the sense that the parameter assumptions have a direct effect on the outcomes of the model. The assumptions are motivated by the theoretical literature and empirical finding, though. Moreover, the model does not allow for an interaction of institutional features or for behavioral adjustments of agents. The empirical analysis that tests the conclusions of the model is conducted in the fourth chapter.

2.2 Empirical aspects of educational production

The empirical estimation of educational production functions is subject to several difficulties. First, the question arises what we are actually concerned with, i.e. what is the main outcome of schooling that we want to examine? This includes the difficulty of measuring the outcome of interest in a comparable way across countries. Second, several problems emerge in the actual estimation of production functions. These problems include limitations of the data such as missing information and missing values, selection bias and endogeneity issues and the choice of the functional form and level of analysis. These empirical aspects of educational production are addressed and possible solutions are provided for the various problems in the following sections.

2.2.1 Measures of education

Despite a multitude of studies on the relationship between educational attainment and labor market outcomes or inputs and outcomes of the educational process, the measurement of skills remains a disputed subject (cf. Borghans et al., 2001). When the measurement of educational outcomes determines the conclusions drawn from empirical studies (e.g. Leuven et al., 2005), the discussion of educational measures is of utmost importance. This section presents different measures of education and discusses their strengths and weaknesses with respect to

the aim of comparing student achievement across countries. It distinguishes between measures of educational levels through formal qualification levels and test scores and provides the motivation why the focus of this study is on student performance data.

Formal qualification levels

Formal degrees are measures of educational attainment such as the highest grade level that has been completed (e.g. primary education, lower secondary) or certificates that are awarded by officially recognized institutions for a certain educational level (e.g. apprenticeship, Bachelor, Master, Ph.D.). Formal degrees have been introduced by public authorities to make education more transparent. Both students and possible employers are given a signal by the formal degree that is awarded to completing a certain education. Recent trends have been to make degrees not only comparable within countries but also between countries. The introduction of the internationally acknowledged bachelor and master degree in many countries is one example; another is the transformation of national degrees into ISCED (International Standard Classification of Education) categories. Formal degrees are the most common measure of education in studies on the relationship between education and labor market outcomes such as wages or unemployment because they are often observed simultaneously with labor market outcomes. However, the meaning of educational degrees differs between countries and even within countries since schooling and other educational systems differ as well. For the estimation of Mincer wage equations as well as for international comparisons, degrees are often transformed into years of schooling. This assumes that each year of education in each educational system has the same benefit for students, though. In secondary education, some countries have central exit examinations to ensure a common standard of high school graduates, while others have not. In tertiary education, the attainment of a certain degree often has less meaning than the university from which it is obtained. Organizations like the “Association to Advance Collegiate Schools of Business” (AACSB) accredit faculties of economics and business nowadays, which leads to a further differentiation of the signal sent to students and employers and increases comparability.

The examples given above show trends towards a higher comparability of formal degrees and a stronger differentiation. So far, formal degrees are hardly comparable across countries, with the exception of internationally oriented higher education programs like business schools. Even within countries, formal degrees are not always set against a certain standard. Therefore, the comparability of formal degrees even within countries is not always ensured and the comparability over time is even more doubtful, especially when standards do not exist or have been shifted over time by educational reforms.

Test scores

Instead of measuring education by degrees or grade levels that have been completed, educational skills can be measured directly by tests. The following paragraphs address the concept of test scores as a measure of skills and discuss different tests. Moreover, the part highlights the economic significance of test scores, i.e. the empirical evidence on the link between tests

scores and economic outcomes. Finally, the student performance studies that are used for the later empirical analysis are introduced.

Tests aim at giving a comparable account of a person's skills and knowledge. The reasons for testing determine the specific test design such as the test content, the test population and the later use of the test results. Internationally conducted tests on student literacy, which are of interest here, have the aim of comparing student performance across countries in order to give countries a relative view on the quality of their schooling systems. The test population normally consists of all students of a specific age or school grade in compulsory education and only a sample of students is tested. The tests cover mostly developed countries but some studies also include larger developing countries. The test content must be either independent of the curriculum or it must be adapted such that students from all participating countries have dealt with the content in a comparable way. The concept of literacy thus seems more suited than testing the knowledge of students. Especially the concept of reading literacy is broadly defined and covers a wide area of useful skills, while the subjects of math and science are rather specific with respect to the test content. International organizations such as the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Co-operation and Development (OECD) are most experienced in conducting these large international comparative studies (e.g. FIMS, FISS, SIMS, SISS, TIMSS, TIMSS-R, PIRLS, PISA).

Longitudinal studies on youths are a different source of student test results. These studies question a sample of a certain cohort at different stages in life. When the cohort of interest is in compulsory education, students are sometimes tested, although the focus of most studies lies rather on health, social behavior or labor market experience. Examples of such studies are the National Longitudinal Survey of Youth starting in 1979 (NLSY), the National Longitudinal Survey of the High School Class of 1972 (NLS72) and High School and Beyond (HS&B) in the U.S. and the 1970 British Cohort Study (BCS70) and the National Child Development Study (NCDS) in Great Britain. These studies also provide detailed background information on students and their parents and have the additional advantage of observing the future educational career and even labor market experience. However, longitudinal studies always suffer from panel attrition and are only conducted within one country, due to their great costs.

Other tests are conducted only within single countries and serve as a means to ensure a certain level of quality at a specific educational level. The most prominent examples are central exit examinations at the end of upper secondary education, which are set and marked external to schools (e.g. *le baccalauréat* in France, A-level degrees in the UK, *Zentralabitur* in some German *Länder*). In some countries like the UK, students are also tested at other key stages of their schooling career. The test results are used to measure school quality (e.g. in UK) or as a selection criterion for university admission (e.g. France, Germany, UK).

Some tests are specifically designed to serve as a signal to universities (e.g. ACT, MCAT, SAT) or prospect employers (e.g. GRE) and have to be taken at the expense of students. They differ

from other student performance tests because participation is voluntary and the tests are conducted mostly by private companies.

The main interest here lies on internationally conducted tests on student literacy, which provide insights into the performance of students in compulsory education and are referred to as student performance tests or studies. These studies have a great value for researchers because background information on students and schools is collected along with test results. The major advantage of well-developed student performance studies is clearly the comparability of tests results and collected data on student and school background across countries. The test and sampling design are chosen such that testing a relatively small sample allows for drawing representative conclusions for the whole target population.

The economic significance of the results such as their direct effects on employment and earnings is less clear than the effects of formal educational degrees because test results and labor market outcomes are not observed simultaneously. Follow-up studies on students tested in PISA 2000 are planned in Denmark but have not yet been realized. Longitudinal studies allow instead for assessing the impact of test scores on employment and earnings for individual countries. Murnane et al. (1995) show that both NLS72 and HS&B test scores have a great effect on wages and that the impact, i.e. the returns to skills, has increased between 1978 and 1986. Follow-up studies show the strong impact of various skills on earnings using the NYLS (Murnane et al., 2001) and of cognitive skills using both NLS72 and HS&B data (Murnane et al., 2000). Lazear (2003) confirms the findings using the National Education Longitudinal Study (NELS) of 1988 and Mulligan (1999) presents evidence from the NLSY. Their evidence for the U.S. suggests that an increase of one standard deviation in math scores at the end of high school leads to about 12 percent higher annual earnings early in the career path. The overall impact of cognitive skills as teenagers on earnings could increase further over the life-time when the returns to experience are higher for more skilled persons and when the impact is increasing over time as indicated by Murnane et al. (1995). Parsons and Bynner (1997) find a strong link between employment and numeracy as measured by the BCS70 for the UK. It is also shown that test scores and years of schooling are highly correlated. Dolton et al. (2005) make use of the NCDS to point out the importance of educational achievement in the career progression.

The International Adult Literacy Survey (IALS) provides information on the direct link between test scores and labor market outcomes by testing adults and collecting data on their labor market outcomes. Although adults are tested and the link between test scores at the age of students and test results in the IALS is unknown, the study shows that skills and literacy as tested by performance studies have a great influence on labor market performance, as could be expected (cf. Green and Riddell, 2003). McIntosh and Vignoles (2001) use both IALS and NCDS data to show a substantial wage return to literacy and numeracy skills in the UK. In contrast to longitudinal studies, the IALS has been conducted in several developed countries and allows hence for a comparison of the link between literacy and earnings across countries. It can be shown that test outcomes are closely linked to labor market success and that distributions of test scores and income are linked over time (Blau and Kahn, 2005). The variation

in the earnings distribution across countries can only partly be explained by differences in the underlying skill distribution, though (Devroye and Freeman, 2001). Bedard and Ferrall (2003) use a larger sample of countries and different sources for skill and wage measures. They find a significant relationship between skills at age 13 and wages later on in life and can link a decrease in the test score dispersion to lower wage inequality.

Besides the effects on individual labor market outcomes, test scores should also affect the economy as a whole. In the theory of endogenous economic growth, human capital is an important factor for explaining labor productivity and hence economic growth. In studies that use years of schooling as the only measure of human capital, economic growth can hardly be linked to human capital accumulation (Prescott, 1998; Hall and Jones, 1999). If instead a combined measure of the quantity and quality of education is used, where quality can be assessed by student performance in international studies (Barro and Lee, 1996; Hanushek and Kimko, 2000), human capital becomes more important for explaining economic growth than does physical capital (Gundlach et al., 2002). Alternatively, Groot and Oosterbeek (1994) show that effective years of schooling are a superior measure of human capital in earnings functions than actual years of schooling. Krueger and Lindahl (2001) establish a significant link between changes in the level of education and economic growth when taking account of measurement error in the years of schooling variable at the country level. The quality of schooling seems to have a higher impact on growth than the quantity of schooling (Barro, 1999; Hanushek and Kimko, 2000).

Two student performance studies are analyzed in the following two chapters: the international study PIRLS on fourth graders and PISA on 15-year-olds. Thirty-five countries participated in the Progress in International Reading Literacy Study (PIRLS). This study was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 2001 and tested fourth grade students (nine- and ten-year-olds) in reading literacy. In the data, extensive information on home and school environments is available through student, parent, teacher and school questionnaires. With 150,000 students tested, PIRLS 2001 is the first in a planned 5-year cycle of international trend studies in reading literacy (Mullis et al., 2003). The data are clustered due to the two-stage stratified sampling design of the study. The schools that participated have been chosen first, before a sample of classes from the targeted grade was drawn. Therefore, the schools are the primary sampling units and not the classes or students.

The Programme for International Student Assessment (PISA) tested 15-year-old students in the subjects mathematics, science and reading proficiency in the first half of 2000. The goal was not to test only the knowledge of students but rather their understanding of the subject matter and ability to apply the acquired knowledge to different situations. The testing was conducted by the OECD throughout its 28 member countries plus Brazil, Latvia, Liechtenstein and the Russian Federation. Apart from test scores, data from student, school and computer questionnaires were collected. These include information on student background, the availability and use of resources as well as the institutional setting at schools (Adams and Wu, 2002). For Germany, a PISA extension study was conducted on behalf of

the states' education ministers, which is called PISA-E (cf. Baumert et al., 2003). This data provides more detailed information on the institutional setting and is matched with the international data for the analysis in Chapter 3. PISA uses a two-stage stratified sampling design as well, which differs slightly from the sample design for PIRLS because the targeted population is not a specific grade but students aged 15. Therefore, schools have been sampled first and then students from the targeted population have been drawn randomly. The sample of countries differs for the two studies.

The scores that are used to measure student performance in PIRLS and PISA are computed according to the item response theory (cf. Hambleton and Swaminathan, 1989). They are the weighted averages of the correct responses to all questions belonging to a certain category, where the difficulty assigned to a question is its weight. Two types of scores are usually reported in student performance studies: weighted likelihood estimates and plausible values. The former score measure is a direct estimate of an individual's proficiency in the respective subject based on his answers and is well-suited to analyze individual level data. The plausible values are drawn from an estimated proficiency distribution. Plausible values are imputed scores based on the students' answers to the test items and provide better estimates at the population level (cf. Mislevy, 1991). The scores have then been standardized to an international mean of 500 and a standard deviation of 100, which facilitates the comparison across countries. The standardization of test scores as the dependent variable in the educational production function has no effect on the regression results with the exception of affecting the absolute size of the estimated effects. Since test scores are an abstract measure and their level has no intuitive interpretation, the absolute size of the effects is irrelevant and only the relative size is important to compare the effects of different input factors.

A main conceptual difference between PIRLS and PISA is the target population that has been sampled. While PIRLS targets a specific grade level (grade four), PISA targets a specific age group (15-year-olds). For a direct comparison of cognitive skills of students across countries, a comparison of age groups like in PISA is more suited. When a representative sample of an age group can be drawn, the mean performance of students can be compared directly without having to take account of the differences in the educational system across countries. The mean performance of students in one grade level instead depends on age at school entry and the rules by which students are promoted to the next grade level. For the analysis of the data, having only one grade level is preferable. Otherwise, one has to account for the selection of students of one age group into different grade levels. Moreover, the sampling over several grade levels is more complicated and therefore less reliable. An additional drawback for the analysis of PISA data is that classes have not been fully sampled because older or younger students do not belong to the target population. The PISA data are hence not well-suited to study e.g. peer group effects.

Conclusion

Formal degrees are most suited to compare educational attainment over time within one country. Moreover, formal degrees are often the only measure of education that is observed

simultaneously with labor market outcomes. Hence, most studies on economic effects of education resort to the measure of formal degrees even for cross-country comparisons. Test scores instead are well-suited for a comparison across countries. However, most international data on test scores are cross-sectional and provide no means to analyze a link between student performance and labor market outcomes. A notable exception is the IALS data. When the focus of research lies on the production of education as in the case of this study, student performance studies provide all necessary information but on prior inputs and are first choice for an international comparison.

2.2.2 Data limitations and estimation problems

The data that are used for the empirical analyses of this paper are student-level data from the PISA 2000 and PIRLS 2001 studies on student achievement. They contain information on student performance measured by standardized tests and on student background, school and teacher characteristics. The latter information originates from questionnaires that are filled in by students, teachers and principals. The detailed description of the data is given in sections 3.2 and 4.2. This section discusses the limitations of the data, the problems that arise in estimating educational production functions and how to overcome these. The main issue is the cross-sectional design of the studies, which do not provide information on prior inputs into the educational production process. Moreover, values for variables are missing because students and teachers did not answer all background questions. In the estimations, the problems of selection bias and endogeneity arise.

Missing information on prior inputs and omitted variables

The educational production function combines all current and prior inputs that are likely to determine educational performance. An empirical educational production function should be specified such that all of these input factors are included. Moreover, the information should enter the production function at the same level as the subject of study, the individual student. Under these assumptions the estimation of the empirical educational production function is supposed to yield unbiased estimators of the effects of the input factors on student performance.

When observational studies lack the information on important input factors or only provide information on current but not on prior inputs, the estimation of educational production functions is limited and the estimators may be biased (Todd and Wolpin, 2003). Under the likely assumption that the missing input factors are correlated with other input factors, the estimated coefficients from these inputs would suffer from omitted variable bias. The problem of omitted variable bias is aggravated when aggregated data is used instead of micro-level data (Hanushek et al., 1996).

A straightforward solution to missing information on prior inputs and some cases of omitted variables is to replace the level model by a value added model by including prior achievement as an explanatory variable. Then, the difference in the achievement levels between two points

in time is explained only by the inputs into the educational production process during this time period. All factors that are time independent and prior inputs can be neglected without risking a bias of the estimator, assuming no long term effects.

The cross-sectional PIRLS and PISA data provide rich information on student achievement, the background of each student, the current school resources including teacher characteristics as well as the institutional setting at the school level. The micro-level data allows for a more precise estimation of coefficients and less bias than aggregated data (Card and Krueger, 1992; 1996). However, no information on prior achievement of students or inputs into the educational production at another time is available. Therefore, a value added model cannot be estimated and the problem of missing information on prior inputs remains. The problem of omitted variables is instead negligible, because the dataset is rich in information on student background and school characteristics and is available at the student level.

As a consequence of the remaining problem of missing information on prior inputs, the production function should be specified in a detailed way and the estimated coefficients must be interpreted cautiously. Besides innate ability, which cannot be measured, student background has been shown to be the most decisive factor in explaining student performance (cf. Hanushek and Luque, 2003; Woessmann, 2003). The student background information includes personal characteristics like sex and information on parents' origin and education. These variables are unlikely to change over time and are hence a good proxy also for prior inputs. Their effect on the cognitive achievement of students can therefore be interpreted as a causal relationship.⁶ Moreover, student background serves as a good proxy for prior achievement of students and a detailed specification of student background can mitigate a possible bias from missing information on prior school inputs (Levacic and Vignoles, 2002).

Variables that describe school inputs and the institutional setting are more likely to change over time. Although students that are tested in PIRLS and PISA are normally at the same school for at least three years prior to the test, some might have changed classes or schools in the past. Teachers and the actual class size may vary from year to year as well. According to Krueger (1999), the effect of class size is strongest in the first year in which a student is placed in a smaller class. When class size does not change strongly over time, this effect should be picked up by level models while value-added estimations may miss the main effect. Overall, current school inputs are only an imperfect proxy for prior inputs. The institutional setting is the framework within which the different actors involved in the production of school education act. The variables describe the distribution of responsibility and other institutional aspects. As institutional reforms take a long time to be implemented, the current institutional setting should accurately describe the setting over the last years, assuming that students stay in the same school. Although one can assume that school inputs and institutions stay fairly stable over time, the estimated coefficients on these inputs should rather be interpreted as conditional correlations and not as causal effects.

6. No distinction is made here between innate ability and student background. Recent evidence suggests that parents' education has little effect on children's education, controlling for the correlation between parents' and children's ability (Black et al., 2005).

Missing values

Missing values for variables from questionnaires are a further problem of observational studies. In many studies, the whole observation (student) is dropped from the regression whenever the value of any explanatory variable is missing. Apart from losing valuable information, dropping students with incomplete answers to the questionnaires leads to a sample selection bias if the values are not missing randomly. Indeed, given that attentive students are more likely to both complete the questionnaire and to answer the test questions, low performing students have a higher probability of being dropped. Thus, dropping observations with missing values leads to an upward bias in test scores, which can be observed for both PIRLS and PISA data for most countries (cf. section 4.2.1).

Instead of dropping observations or replacing missing values arbitrarily by other values, the imputation of missing values is a more sophisticated approach to overcome the problem of missing values and to avoid a sample selection bias. Missing values can be predicted on the basis of regressions of the variables with missing values on a certain selection of student background variables that are available for almost all students. Students who did not answer these elementary background questions or did not complete the test have to be excluded from the sample. Students who have many missing values have also to be excluded because the whole information on the observation is not reliable any more when too many values are predicted and not observed. For the prediction of missing values, linear regression models can be used for continuous variables and probit and ordered probit models for qualitative variables. It has to be assumed that the estimated coefficients are also representative for the observations which cannot be included in the regressions because their missing values are to be predicted. The specification of the regressions used for the prediction should be chosen such that the variation of the dependent variable can be well explained and that the variation in the predicted values is high. To this end, a parsimonious but well chosen set of explanatory variables is preferable to including all available variables. In addition, the standard errors of estimations using imputed values can be adjusted for the uncertainty of the imputation process and the share of missing data (Schafer and Schenker, 2000). In the empirical applications of this study, results for the data both with and without imputed values will be presented instead of adjusting the standard errors for the imputation process. Fuchs and Woessmann (2006) have shown that the adjustment hardly affects the results.

The prediction of missing values on the basis of regression results is clearly no impeccable solution and has several shortcomings. First, the limited set of variables that is available for predicting the missing values has a limited explanatory power and is often ill-suited to explain other background variables. Second, the variation of the variables decreases because the variation of imputed values is always lower than the 'true' variation. Third, it is doubtful in how far the observations with missing and without missing values are truly comparable with respect to unobservable characteristics and whether the estimated coefficients are thus representative for both groups. In spite of the shortcomings of the method of imputation, it is by far preferable to dropping observations or arbitrarily replacing missing values. With the right specification of the regressions used for the prediction, the imputed values vary greatly as well and the

information of the non-imputed values of the observation is kept. The strongest argument for the imputation of missing values is the risk of a sample selection bias that arises otherwise. A recent application of sophisticated imputation methods to student performance data shows that imputation is unlikely to introduce an additional bias and that the assumption of conditionally randomly missing data does not affect the results (Fuchs und Woessmann, 2006). Due to the shortcomings of the method, the estimations should additionally be performed for the original data only and be compared to the results for the imputed data as a robustness check. Any inferences that are merely based on the imputed data and not on the original data should then be interpreted carefully as they seem fragile (Mukherjee et al., 1998).

Selection bias and endogeneity

In the estimation of empirical educational production functions the problem of endogenous explanatory variables arises. This concerns mainly school-level variables such as teacher characteristics and resources. Endogenous variables are correlated with the error term of the regression equation by definition. There are two sources of endogeneity in this context: omitted variables and simultaneity. Omitted variables like student's innate ability have to be correlated with the endogenous explanatory variable in order to bias the estimator. When an endogenous variable is determined simultaneously with the dependent variable, we speak of simultaneity. Often, these two sources of endogeneity are hardly distinguishable from each other. For the estimation of educational production functions, the non-random allocation of students into schools and classes is the main reason why school-level variables cannot be considered as exogenous to student performance. Students can choose between schools and school types in many countries or are allocated into certain schools or classes by teachers and principals. The degree of selection between schools and classes depends on the school system and thus differs between countries (cf. Ammermueller et al, 2005; Woessmann and West, 2006). Alternatively, principals and administrators can select teachers or determine the level of resources according to student achievement. This reverses causality and complicates the interpretation of regression results. A prime example is the often cited positive link between class size and student achievement in production functions estimated by OLS (cf. Woessmann and West, 2006). Low performing students seem to be allocated to smaller classes in most countries.

There are several ways to avoid or overcome a selection bias. First, instead of using data from observational studies, researchers can revert to experimental data. In an experiment, one can control for any selection and allocate students randomly to schools and classes. This was the main objective of the project STAR (cf. Krueger, 1999; 2003). However, as in any social experiment where stakes are high, parents tried to place their children in the best classes or decided to abandon the experiment. Moreover, people who participate in a social experiment may act differently because they are being evaluated.

Second, the reduced form of the empirical educational production function can be replaced by a full structural model. In a structural model, the selection of students and the allocation of resources are modeled explicitly. Then, the production function and a separate equation

for each endogenous variable are estimated simultaneously and the coefficients can be interpreted as causal effects. In a structural model, each equation has to be uniquely identified, though. This implies that the researcher has to select at least one variable which determines the endogenous variable of choice but is not related to either student achievement or to any other endogenous variable. When student achievement and resource allocation are closely linked, it is doubtful whether the identification restrictions can be reasonably fulfilled.

Third, selection between school types and schools can be controlled for by including school type as an explanatory variable or by estimating school-fixed effect models, respectively. Any unobservable differences between school types and schools can then be neglected. However, school-fixed effect models eliminate any variation at the school level and hence allow only for an estimation of effects for variables that vary below the school level.

The fourth and most prominent way to overcome the selection bias is the use of instrumental variables. These have to be correlated with the endogenous variable and uncorrelated with the error term, i.e. they are linked to student achievement only indirectly via the endogenous variable that is to be instrumented. Examples of instrumental variables for class size are student enrollment at the given grade (Akerhielm, 1995), average class size at the grade level (Ammermueller et al., 2005; Woessmann and West, 2006), rules on maximum class size (Angrist and Lavy, 1999) and population variation (Hoxby, 2000). Besides the question of the validity of any instrument, the problem of the interpretation of the estimated coefficient arises. When instrumental variables are used in a simple two-step least squares estimation, the coefficient is only representative for that sub-set of the population for which the instrument actually matters. In terms of the evaluation literature, this is a so-called local average treatment effect (LATE) interpretation (cf. Carneiro et al., 2003). When a rule on the maximum class size is used as an instrument for example, the coefficient is representative only for those students who are in classes close to the maximum class size to which this rule might apply. Therefore, instrumental variables have to be chosen carefully and the interpretation of the results has to be stated explicitly.

The description of the various ways of selection that arise in the process of education and the multitude of methods to deal with these problems underlines that there is no standard solution to this problem. The choice of method depends on the problem statement, the kind of selection that arises and the structure and richness of the available data. The empirical analyses of this study use very detailed information of student background, test for selection on observables and control for school types and employ school-fixed effects models in addition.

Level of analysis and functional form

The level of analysis and the functional form of empirical educational production functions greatly depend on the available data. The level of analysis ranges from micro-level student data to aggregated country-level data. The choice of the level of analysis also depends on the problem statement and thus the outcome variable of interest and the degree of variation that is observed in both the outcome and explanatory variables. Micro-level studies have several

advantages compared to studies relying on aggregated data. First, the number of observations and the variation in the data are higher in micro-level studies. Second, aggregated data are more prone to measurement error and hence more likely to yield biased estimators (Goldstein, 1995; Hanushek et al., 1996). Therefore, the data should enter the educational production function at the lowest level that is available, which is mostly the student level.

The empirical specification and functional form of the educational production function should be carefully chosen with respect to the problem statement. When student achievement is to be explained in general, all input factors that might affect educational production ought to be included in the empirical production function. When the focus of research lies instead on the effect of student background, the inclusion of other input factors might only confound the findings and a reduced form of the production function is more appropriate. Regarding the functional form, linear models are chosen in most empirical studies (cf. Levacic and Vignoles, 2002) and are also used for the estimations in this study. This choice of form facilitates the estimation by linear regression models such as OLS. However, the data should be carefully tested on non-linear effects for which a different form such as a trans-log functional form is more appropriate. A linear form is restrictive and is rejected by the data in some instances (cf. Figlio, 1999). A flexible functional form can also be achieved when categorical variables are transformed into dummies, i.e. the effect is estimated for each level of a variable separately instead of imposing one effect for all levels. Moreover, interactions between variables can be introduced to allow for non-linear effects.

A further issue with respect to the estimation of production functions is the structure of the data. Standard econometric techniques such as OLS assume that the error terms between observations are not correlated. Hierarchical data such as most student performance studies violate this assumption. In these data, individual observations (students) are not independent from each other because they have not been chosen randomly among all students in a country. Instead, schools or classes have been chosen as the primary sampling units (PSU) and then students within these PSU have been selected (cf. sections 3.2 and 4.2 for a detailed description of the data). It is hence likely that students from the same PSU share some unobserved characteristics, which leads to a correlation of their error terms in the empirical production function. Ignoring the clustered sampling design of the studies in the estimations can lead to spurious regressions and underestimated standard errors (Moulton, 1986; 1990). Multilevel models or other econometric techniques that adjust standard errors have to be employed instead of OLS to estimate production functions with hierarchical data. While cluster-robust regressions (cf. White, 1980, p. 821; Deaton, 1997, pp. 73-78), also known as survey regressions, are robust by allowing for any dependence of observations within the PSU, multilevel models have been criticized for being unnecessarily complicated although the results are similar to OLS estimates (de Leeuw and Kreft, 1995). However, under certain assumptions feasible generalized least squares (FGLS) leads to more efficient estimates. FGLS controls for heteroskedasticity more efficiently but needs stronger assumptions on the correlation of explanatory variables and error terms to be consistent compared to CRLR, which is completely robust (see Wooldridge, 2002, pp. 153-160). Alternatively, production functions can be estimated by quantile regressions to uncover any differences in the estimates

for different quantiles of the conditional distribution. First evidence suggests that the difference in the estimates of resource or institutional variables is hardly ever significant between quantiles (e.g. Ammermueller, 2006; Woessmann, 2005). However, for the estimation of peer effects quantile regressions provide additional insights into the production of education in class (e.g. Rangvid, 2003).

In a review on difficulties encountered in the estimation of educational production functions, Levacic and Vignoles (2002) summarize the criteria for high quality studies. The first criterion is that micro- or student-level data with detailed information on prior achievement, student background and school characteristics is used for the estimation. Second, the possibility of endogenous variables should be addressed and alternative estimation methods such as instrumental variables should be used. Third, the assumptions that are imposed on the data by a specific functional form should be tested and the form adapted to the findings. Finally, the estimator should account for the structure of the data. It should be added that missing values are correctly dealt with and that robustness checks are run on the data. The limitations of the data and the consequences for their empirical analysis are considered in the analysis of the data in the following two empirical chapters.

3

Explaining the difference in student performance between countries

This chapter addresses the question how the effects of student background, resources and institutional features on student performance as measured by test scores differ between two countries with very different schooling systems.¹ The Finnish system with a high degree of school autonomy and a single school type is compared to the German system with a highly differentiated schooling system. The effects of different inputs into the educational process are estimated separately for the two countries and their contribution to the difference in test scores is assessed by a decomposition of the test score gap. This approach only allows for estimating the effects of variations of the institutional setting within a country. Therefore, the focus is not on schooling institutions at a national level, but rather on a comparison of the effects of local variations in institutions and other inputs, given the institutional framework of the country.

3.1 Introduction

The publication of the PISA (Programme for International Student Assessment) outcomes showed that test scores for reading proficiency of Finnish students of age 15 are the highest in Europe and that those for mathematics and science rank third and fourth, respectively. The test scores of German students of comparable age ranked well below the OECD average in all three test subjects (Baumert et al., 2001; OECD, 2001). The reasons underlying the great difference in student test scores between the two countries in PISA are the subject of this analysis.

Differences in PISA performance concern the level of the average test scores but also the dispersion of the score distribution. The first question that arises is which factors can explain the different results in the two countries. Are these non-institutional factors like student and school characteristics or is it rather the institutional setting that makes the difference? The second question is how certain factors contribute to the difference in student test scores.

1. I would like to thank Charlotte Lauer, Denis Beninger, Peter Jacobebbinghaus, François Laisney, Axel Pluenecke, Manfred Weiss, two anonymous referees as well participants at the EALE conference in Lisbon for constructive comments on this chapter. The article Ammermueller (2006) is based on this chapter.

Are the characteristics more favorable in the high scoring country or are the returns to these characteristics in terms of test scores more advantageous? For example, the low German test scores could be due to a higher share of students from a lower social background, as measured by parent's education, in Germany than in Finland. However, if the assumed negative effect of a lower social background were smaller in Germany, the overall impact on the average test score could still be comparable to the impact in Finland. The available resources and the institutional setting of schools might explain the difference in test scores as well. Therefore, the paper examines the differences between the test score distributions in Finland and Germany and decomposes them in order to quantify the different effects. Thus, the analysis aims at disclosing possible sources of the low test scores in Germany and thereby gives guidelines where improvements of the schooling system are most feasible. Finally, these countries allow for comparing a diversified (Germany) to a single type schooling system (Finland).

Although there is an extensive body of literature on student performance in Germany, few studies consider the aspects of educational production or provide an in-depth country comparison. Gundlach et al. (2001) discuss the productivity of schools across European countries but provide no detailed country comparison. Related studies mainly focus on the bivariate correlation between inputs and test scores or use students' basic cognitive abilities to explain test scores (Baumert et al., 2001; Artelt et al., 2001). Another study that employs multivariate methods uses unprocessed data that ignore the problem of missing values and include no information on school types (Fertig, 2003). Dropping students with missing information for some variables might lead to a sample selection bias and neglects the use of the entire set of information that is available for the analysis. Ignoring school types in an analysis of the diversified German schooling system likewise ignores potentially relevant information. The regression analysis conducted in this paper uses a unique dataset with imputed data for missing values and school type information from an additional data source.

The main results are that the measurable characteristics of German students and schools cannot explain the test score gap, except for the lowest deciles of the test score distribution. Instead, the use of resources and the transformation of student and school characteristics are more effective in Finland. Test scores of students depend to a higher degree on student background, as measured by parent's education and the number of books at home, in Germany than in Finland, which explains the higher inequality in test scores in Germany. The results might imply that streaming in Germany affects students in lower school types negatively and leads to greater inequality of educational achievement. It remains unclear in this study, however, if this can be attributed to the effect of school types per se or student background and innate ability that determine the allocation process of students into school types.

The remainder of the chapter is structured as follows. The second section introduces the dataset and describes the data for the two countries of interest. The third section discusses the determinants of educational performance as measured by test scores. In the fourth section, the Oaxaca-Blinder and Juhn-Murphy-Pierce decompositions are performed. Finally, the fifth section concludes with a summary of the findings and their political implications.

3.2 The dataset

The analysis builds on the international PISA data which was briefly introduced in section 2.2.1. A sample of 15-year-old students was tested in the subjects reading and mathematics or science proficiency in the first half of 2000. The publicly available data for the countries Finland and Germany are being used. For Germany, additional student-level information on school types is taken from an extended version of the PISA study.² The two data sources are merged at the student level and then the information is extracted. For a detailed description of the German PISA study see Baumert et al. (2001). Välijärvi et al. (2002) present an analysis of the Finnish results while Gautschi (2005) provides an overview on the Finnish schooling system.

The student test scores are standardized to an international mean of 500 and standard deviation of 100, which facilitates the comparison across countries. In the analysis, the weighted likelihood estimates are chosen as the outcome measure because they estimate an individual's proficiency in the respective subject (cf. section 2.2.1). The values given for the population parameters might slightly differ from other publications (e.g. OECD, 2001), which use plausible values that are drawn from an estimated ability distribution and provide better estimates at the population level. The means and standard deviations of the scores and variables used in the analysis are presented in Table A3.1 in the appendix and are weighted by the students' sampling probability. Table A3.2 displays statistics of selected variables separately for the different school types in Germany. The means of the variables show how greatly the characteristics of students and schools vary between school types. This reinforces the observation that relevant variables are omitted by excluding information on school types. The standard deviations imply that the variation within school types is also high, except for the school being public or not, but smaller than the overall variation.

The sample for Finland comprises 4,855 students in 154 schools that participated in PISA 2000 and completed a reading proficiency and mathematics or science test. In Germany, 4,917 students in 219 schools are in the sample. Together with the background information that is provided, the PISA data are one of the most recent and detailed publicly available micro-level data on student performance for the two countries and are internationally comparable. The data are clustered due to the stratified sampling design of the study. The schools that participated have been chosen randomly in a first step, before a random sample of the student target population was drawn. Therefore, the schools and not the students are the primary sampling units.

Section 2.2.1 discussed the benefits and drawbacks of targeting an age group of students. The sampled 15-year-old students are either in grade eight or nine in Finland and in grade eight, nine or ten in Germany. Table A3.3 displays the average age of students in months and the average reading test score by grade. Students in higher grades are slightly older and

2. In Germany, an extended version of the PISA study was conducted on behalf of the states' education ministers. However, the so-called PISA-E data is not well-suited for a comparison to the Finnish data due to the huge difference in sample size and missing information in the publicly available data-files.

score higher than students in lower grades. While students in Finland are almost always promoted to the next grade level each year, students in Germany have to perform above a certain threshold. Otherwise they must repeat the grade. Due to this strict rule 2.98 percent of all German students had to repeat a class in the school year 2002/2003 (KMK, 2004). Due to the difference in promotion rules, the selection into grade levels differs between the two countries. While the grade level of a student should be determined only by his age in Finland, it depends on both age and previous performance in Germany. Therefore, students in lower grades are likely to be a negative selection of students in Germany but not in Finland. In order to check for a possible bias of the results, the analysis is also performed for a reduced sample, which consists only of students in grade nine in both countries. This sample should be less affected by a possible selection due to promotion rules.

Missing values for student and school background variables are a main problem of the data as described in section 2.2.2. For Germany and Finland, up to 16 percent of the values of key variables such as parents' education are missing. Test scores are reported for all students. Table A3.4 presents the percentage of missing values. Commonly, the whole observation (student) is dropped from the regression whenever the value of any explanatory variable is missing. This would lead to a decrease of observations of 43 percent in Finland and 60 percent in Germany. Apart from losing valuable information, dropping students with incomplete answers to the questionnaires leads to a sample selection bias if attentive students are more likely to both complete questionnaires and to answer the test questions. Thus, dropping observations with missing values would lead to an upward bias in the test scores of the estimation sample. Table A3.1 displays a small upward bias in test scores for Germany but hardly any change for Finland. The standard deviation of test scores decreases as well for Germany but not for Finland when students with missing values are dropped.

The approach chosen here to overcome the problem of missing data is to predict missing values on the basis of regressions on those background variables that are available for most students like age, sex and the grade a student is in (see section 2.2.2 also for assumptions and shortcomings of the imputation method).³ Students who did not answer these elementary background questions or did not complete the tests are excluded from the regressions, as well as students with more than 10 missing values. This applies to less than one percent of the sample but leads to a significant increase in mean test scores and a lower standard deviation in Germany. The exclusion of outliers is necessary so that the analysis is not dominated by a small and unrepresentative sub-sample of the student population.⁴ The descriptive statistics and the regression results are also given for the original data without imputed values in Tables A3.1 and A3.4, respectively, where all students with at least one missing value are dropped. For the following analyses the sample with imputed values is used if not stated otherwise.

3. This method is comparable to the STATA *impute* command, which uses only linear regressions (cf. Goldstein, 1996) and has been employed in several recent studies using student performance data (e.g. Ammermueller et al., 2005; Woessmann, 2003; 2005).

4. Moreover, students with an unrealistically low score of below 200 points (26 students), students from one school in Finland with identical test scores (5 students) and students not in grade 8, 9 or 10 (65 students) were dropped from the regressions as outliers.

3.2.1 *Distribution of test scores*

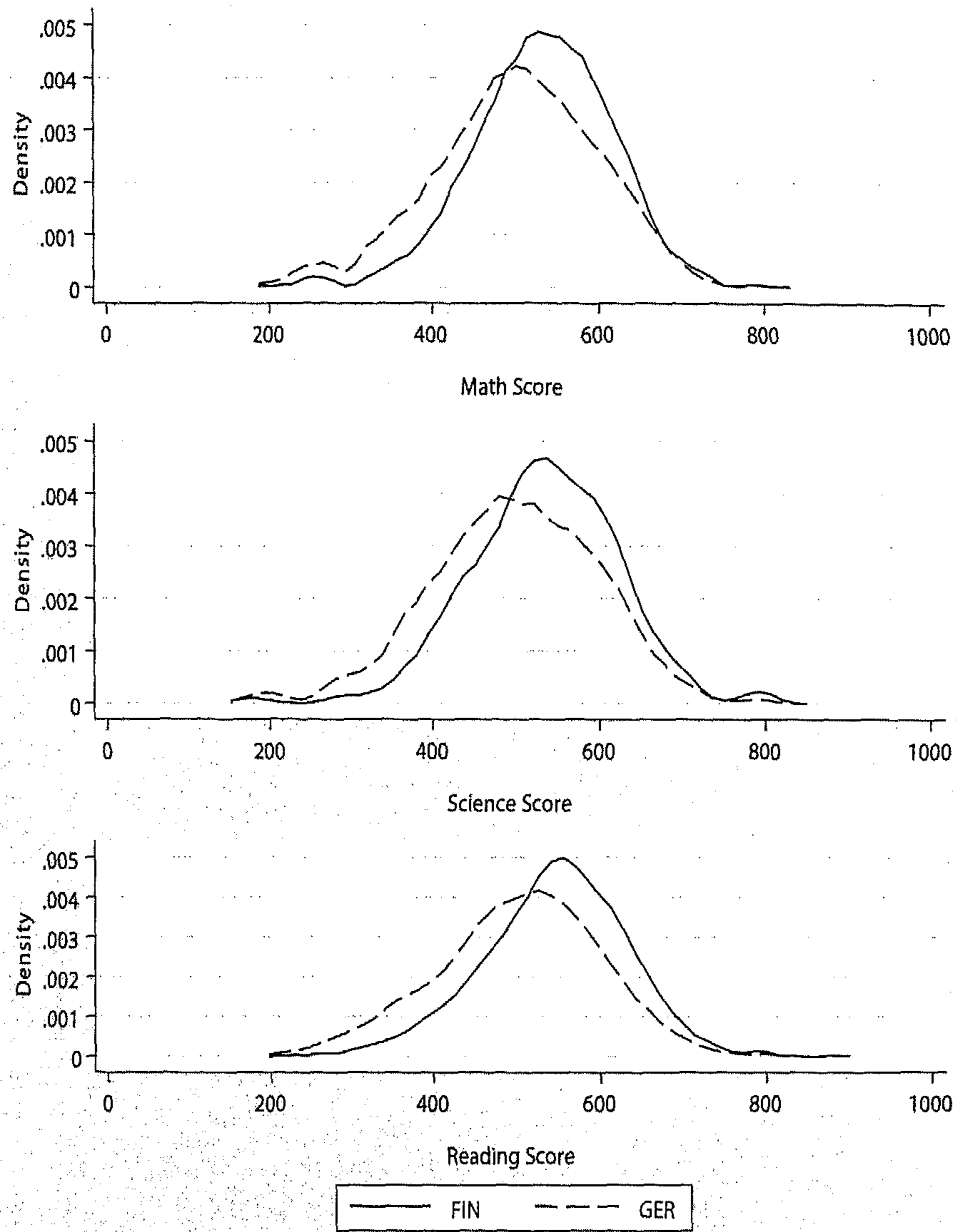
In this part, the distributions of test scores for Finland and Germany are presented graphically. For each subject, non-parametric kernel density estimates describe the score distribution of the two countries.⁵

Figure 3.1 displays the test score distributions for the three subjects for both Finland (FIN) and Germany (GER). The Finnish scores are on average higher than the German scores, which can be seen in the position of the Finnish distribution that is to the right of the German distribution and the higher weighted average score. The average test scores and their standard deviations as well as the minimum and maximum value for each variable are presented in Table A3.1 in the appendix. The mode of the Finnish distribution is also clearly to the right of the German distribution. Moreover, Finland has not only a higher share of students with high test scores but especially fewer students with low and very low test scores than Germany, which has a relatively fat left tail. Despite the higher average scores, Finland has a lower standard deviation of scores. This pattern holds for all three subjects in which the students have been tested.

Finland exhibits the following characteristics in its test score distributions compared to Germany: higher average scores, a higher mode, a smaller variation of test scores and especially fewer students with very low test scores in all subjects tested. The question therefore arises if the reasons for this great difference in test scores between Finnish and German students can be identified. The subsequent analysis focuses on the reading proficiency of students because the respective test scores are available for all students. The distributions of the other test scores suggest no important difference between the three subjects, in which the students were tested. The following analyses have also been conducted for the subjects of math and science. The results are available with the author but are not reported here because they support the conclusions and differ only in the size of some effects.

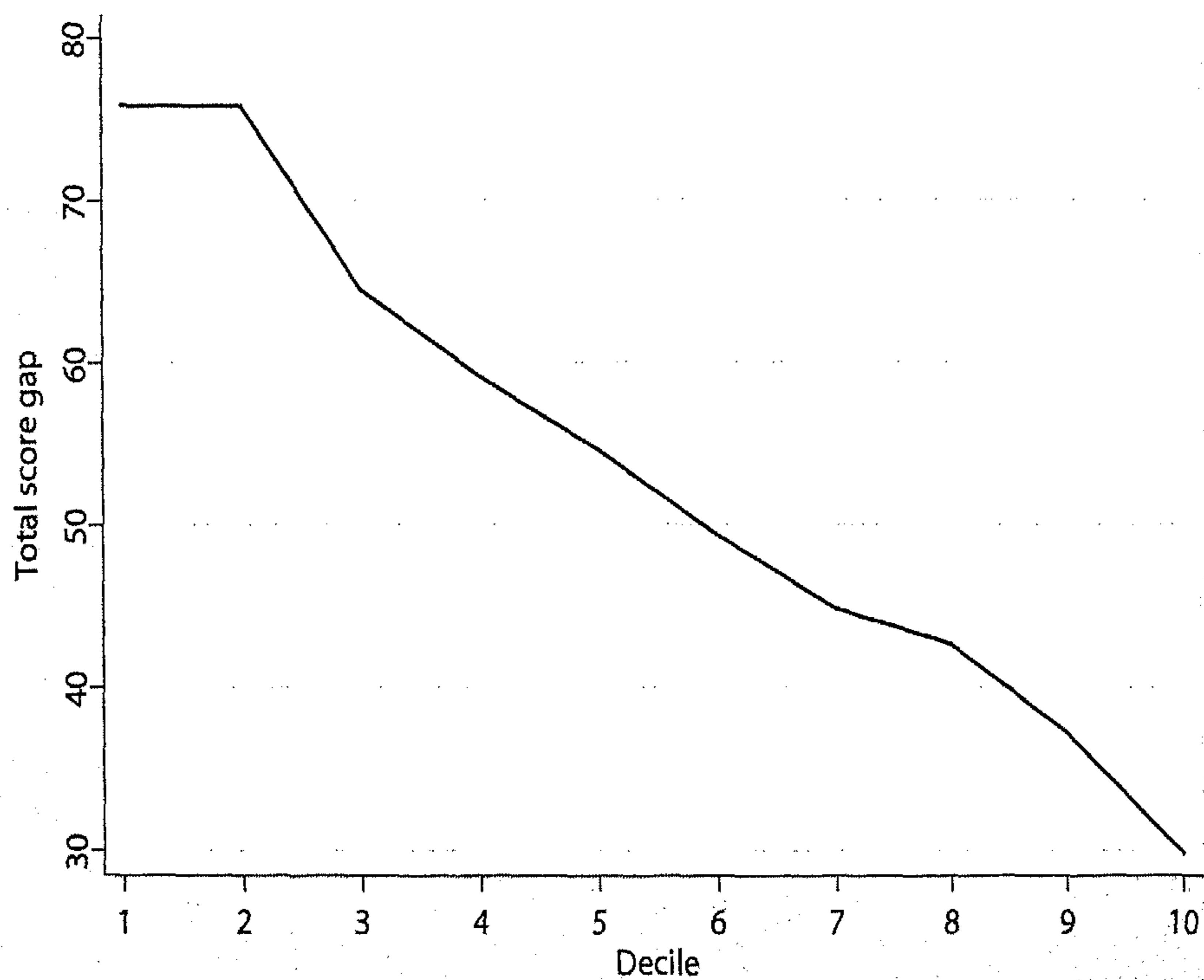
5. The bandwidth was chosen using Silverman's rule of thumb. For a description of density estimation see Silverman (1986). The kernel density estimates are preferred to histograms to show the differences between the distributions continuously. Moreover, the decomposition method in section 3.4.2 also relies on density estimates.

Figure 3.1
Test score distributions



The total reading score gap between Finland and Germany is shown in Figure 3.2. Students' sampling probabilities are taken into account when dividing the test score distribution into deciles. The gap averages 54 points but is declining strongly along the deciles of the distribution. While it is 75 points for the lowest scoring decile of students, it reduces to 30 points for the highest scoring 10 percent of students. The difference in test scores is significant at each decile. In particular the relatively low scores of the lower part of the German test score distribution seems to be mainly responsible for the low average score compared to Finland. However, also the highest scoring German students do not attain the same level as the Finnish students. The inequality in the test score distribution is hence much higher in Germany than in Finland. In the next step, the factors that affect the test scores in either country are analyzed.

Figure 3.2
Total reading score gap



3.3 Determinants of reading proficiency scores

3.3.1 The educational production function

In order to assess the importance of the various input factors for the process of educational production in the two countries, educational production functions are being estimated. Section 2.2.2 outlined the limitations of the cross-sectional PISA data and further estimation problems. The following model of an educational production function is being estimated separately for each country using cluster-robust linear regressions (White, 1980, p. 821; Deaton, 1997, pp. 73-78; section 2.2.2). The specification of the model builds upon the theoretical model developed in section 2.1.1 but is mainly inspired by the information provided in the PISA data. Inputs like prior resources and student ability have to be neglected in the following model (cf. Todd and Wolpin, 2003):

$$(3.1) \quad T_{is} = \beta_0 + G_{is}\beta_1 + B_{is}\beta_2 + R_s\beta_3 + I_s\beta_4 + S_s\beta_5 + v_s + \varepsilon_{is},$$

where T_{is} is the reading test score of student i in school s , G_{is} comprises grade level dummies and a variable on student's age in months, B_{is} is a set of student background variables, R_s is a set of variables on school resources, I_s represents institutional variables, S_s school type variables and v_s and ε_{is} are the error terms at the school and student level. The groups of parameters β_0 to β_5 are to be estimated. Table A3.1 in the appendix presents all explanatory variables and their descriptive statistics.

The dummies for grades eight and ten estimate the advances that have been made by students in reading literacy compared to grade nine, which is the reference group.⁶ In Finland, students are either in grade eight or nine while in Germany some students are also in grade ten. The continuous variable age is measured in months and captures also the effect of grade repetition. German students are on average slightly older and in a higher grade than Finnish students.

Besides innate ability, student background has been shown to be a decisive factor in explaining student performance (cf. Hanushek and Luque, 2003; Woessmann, 2003). The background B_{is} includes personal characteristics like sex and information on parents' origin and education. These variables are unlikely to change over time and may hence serve as a good proxy for prior inputs. The estimator of their effect on the cognitive achievement of students should therefore not be biased. However, β_2 underestimates the total effect of student background on student test scores if there is an indirect effect via the school types S_s . This is the case when the allocation of students to schools does not only depend on innate ability but also on parental background. Therefore, the coefficients of student background should better be interpreted as the lowest boundary of the effects, especially for Germany with its several schooling types.

6. The estimates are likely to be biased upward because the group of 15-year-old students in the upper grade is likely to be a positive selection while those in the lower grade a negative selection of the students in the respective grade. Also refer to the discussion of the age and grade effect in sections 2.2.1, 3.2 and 3.4.2.

The current resources R_i describe parts of the schooling system that depend mainly on financial investments from the public side. The teacher-student ratio at the school level is used to measure the input of teachers for each student. Instead, actual class size would have the advantage of measuring this input more directly, but the estimator of class size is likely to be biased. For the class size estimate, selection of students within schools adds to the problem of selection between schools (cf. Woessmann and West, 2006). Indeed, low performing students might be put in smaller classes in order to foster their learning. School type dummies can only control for selection between schools. Under the assumption that students do not switch schools and that the teacher-student ratio is roughly constant over time, the current ratio is a reasonably good proxy for teacher input per student over the last years. The problem of endogeneity is discussed in more detail at the end of this section.

The institutional setting is the framework within which the different actors involved in schools act. It may affect the motivation and incentives, especially of students and teachers (cf. Bishop and Woessmann, 2004). The variables describe the distribution of responsibility and other institutional aspects. As institutional reforms take a long time to be implemented, the current institutional setting should accurately describe the setting over the last years, assuming that students stay in the same school. In Finland, students usually stay in the same comprehensive school over the entire period of compulsory education of nine years while German students commonly change school after four years of elementary school.⁷ Given that the tested students have already spent four to five years at their secondary school, the effect of the former elementary school should be negligible. Under the mentioned assumptions, the estimators for resources and the institutional setting of the school should not be biased, especially since controls for school types are included but should be interpreted cautiously.

Finally, the school type variables indicate the type of schools a student attends. This can be five types in Germany and only one type in Finland. German students are allocated to secondary school types after their fourth school year according to their performance in elementary school.⁸ Assuming that innate ability of students in the fourth and ninth school year is not independent, there is a problem of endogeneity between school type and student test scores because both are determined by innate ability. Moreover, as educational performance in elementary school, the preference for school types and thus the allocation to a school type are also determined by student background, the school type coefficient might include a part of the student background effect on student achievement. Hence, the school type coefficient consists of the 'true' school type effect, an effect of sorting by innate ability and an additional impact of student background on student test scores via school type. The coefficient β_5 can therefore only be interpreted as a partial correlation.

Note that the specification of the production function is chosen in such a way that all coefficients of the same category of variables are expected to have the same sign. Student background variables are expected to have negative coefficients while resource and institutional

7. Information on the educational systems is taken from Eurybase (2003).

8. Teachers at elementary school write recommendations for each student, and then parents have to apply at schools. Only the degree from the higher secondary school (Gymnasium) allows attending university. The vocational school (Berufsschule) is for students in an apprenticeship.

variables are expected to have positive coefficients. This is important for the identification of the effects in the decomposition in the following sections. Only for the grade level, grade nine and not grade ten has been chosen as a reference category because there are no students in grade ten in Finland.

Selection might not only take place between school types but also between schools of the same school type and within schools. Examples include the selection to specific schools or selection between classes or grade levels within schools. When selection by unobservable characteristics occurs and school resources and institutions differ systematically for students with a different level of e.g. innate ability, the class and school level variables in equation (3.1) would be endogenous and the estimators would be biased. This could concern the estimations for both Finland and Germany.

In Germany, the selection of students into school types is extremely strong as shown in Table A3.2. Selection between schools of the same school type is rather unlikely. Although there is free choice of schools and schools are free of charge, the nearest school of the appropriate school type is chosen normally. Schools are mainly financed by the state (80 percent) and to some degree by the municipality and local district (20 percent) in Germany (cf. BLK, 2003; Klös and Weiß, 2003). Therefore, school resources do not depend on the neighborhood as in other schooling systems (e.g. the U.S.).⁹

For Finland, similar arguments apply. Students mostly attend the nearest school. Due to the rural character of the country, selection between schools would only be possible in the population centers of Southern Finland but parents still tend to choose the nearest school (cf. Gautschi, 2005). From age seven, when Finnish children enter school, they stay within the same class till the end of primary school after grade six. The following three years of compulsory education, Finnish students attend upper school. Only thereafter, students either enter college preparatory school or a vocational school (Gautschi, 2005). Hence, selection by performance takes place the first time at around age 16 for Finish students.¹⁰

3.3.2 *The estimated effects*

The effect of the characteristics on student test scores is estimated in a regression of the individual student test score on the explanatory variables separately for Finland and Germany

9. Running survey regressions of the student background variables books at home and parents' education on school resources and institutions shows that school characteristics are not related to student background, once controls for school type are added. This shows that no selection on observable characteristics takes place. However, selection on unobservable student characteristics cannot be tested. Selection within schools is unlikely because students have already been selected to school types. Moreover, the variables of concern are determined at the school level, not at the class level, and should hence not be affected by selection within school.

10. Running the same regression as for Germany suggests that the school level variable on whether schools are allowed to select their students is related to observable and maybe also to unobservable student characteristics. The further analysis takes into account the possible endogeneity of the student selection variable.

(see equation (3.1)).¹¹ Due to the clustered design of the PISA data, cluster-robust linear regressions, also known as survey regressions, are used for the estimations (cf. White, 1980, p. 821; Deaton, 1997, pp. 73-78; section 2.2.2). These correct the standard errors for the clustered data design, implying an interdependence of error terms between students within the same school. As students from different schools and countries have different sampling probabilities, the sampling weights available in the data are used for the estimation of equation (3.1). The outcomes of the weighted survey regressions with the dependent variable reading proficiency score are presented in Table A3.4. Including the imputed values for the estimations does not affect the qualitative interpretation of the results but makes them more representative of the student population.

The R^2 of the regressions indicates that more than half of the variation in the German test scores can be explained but only 17 percent of the Finnish variation. The test scores of students in Germany depend therefore more highly on factors that can be observed and far less on innate ability and other unobserved factors than in Finland. Even when the school type variables are not included in the regression for Germany, the R^2 of .34 is still twice as large as for Finland. Excluding the age and grade level variables in addition leads to an R^2 of .15 in Finland and .34 in Germany.

An additional year of schooling adds between 38 points in Germany and 48 points in Finland to the test score, which is almost half a standard deviation. However, the estimates are likely to be overestimated due to the selection of 15-year-old students in the respective grade. This magnitude underlines how large the test score gap of 54 points between Germany and Finland actually is. The coefficient for tenth grade can only be estimated for Germany because there are no students in grade ten in Finland. The coefficient for age is negative and much larger in absolute terms in Germany, which might be due to the higher share of repeat students in Germany as compared to Finland.

The student background coefficients are highly significant and have a high impact on student test scores.¹² For example, students whose parents do not even have completed secondary education score 35 points lower in Germany, respectively 26 points lower in Finland, compared to students whose parents have completed tertiary education, all else equal. The negative impact of an unfavorable student background is higher in Germany than in Finland for this example. Girls perform significantly better than boys in the reading literacy test, in particular in Finland. Students who were born abroad or whose parents immigrated score lower than comparable non-immigrated students, especially in Finland where the share of these students is only three percent compared to 20 percent in Germany. The number of books at home has a highly significant and large effect on test scores.

11. Characteristics here imply all measurable characteristics, including the grade level, student background, resources, institutional setting and school types.

12. This is confirmed by the marginal effects and is also the result of studies using TIMSS data (e.g. Ammermueller et al., 2005).

Due to the large share of migrant students, i.e. students of which at least one parent was born abroad, in Germany compared to Finland, the question arises whether German students can be treated as a homogeneous group. When the production functions are estimated separately for migrants and non-migrants in Germany, the regression results show that the coefficients differ only for two variables between the two groups of students at the five percent-significance-level (cf. Ammermueller, 2007). The variables are “no secondary education” and “gymnasium” and the effects are higher for migrant students. Furthermore, students in schools with a large share of migrants might be affected by peer effects. This would be the case when students have no difficulty with the German language themselves but learn less because many of their peer students have low German language skills. The inclusion of a peer variable that measures the share of migrant students in the student sample at the school level in the production function (equation (3.1)) could indicate whether there are negative externalities from a high share of migrant students on other students. The peer variable is significantly negative when only student background, school resource and institutional variables are included. Adding the school type dummies to the production function leads to a positive and non-significant coefficient for the share of migrant students in school, however. This can be explained by the fact that the share of migrant students varies strongly between school types. While it is below 14 percent in the highest school type (Gymnasium), it is around 40 percent in the lowest school type (Hauptschule). Possible peer effects of the student composition are hence already included in the school type dummies in equation (3.1).

The coefficients for resource variables are never significant. The teacher-student ratio and having no lack of material are not significantly related to test scores. A high share of teachers with tertiary education leads to a non-significant increase in test scores in Germany but not in Finland. The coefficient for instruction time cannot be estimated for Finland due to the lack of variation in the variable. The variables describing the institutional setting are not significant except for the power of schools to select their students in Germany. The variation of the institutional setting within countries is not very large however, so that inter-country comparisons are more suited for analyzing their effects (cf. Chapter 4). School types exhibit highly significant effects in Germany in reference to comprehensive schools. Students who attend a low secondary school score 34 points lower than comparable students in comprehensive school in Germany, while students in higher secondary schools score 85 points higher.

After having shown the determinants of test scores in reading literacy in the two countries, the following section compares the results more systematically by decomposing the score gap between Finland and Germany into different effects.

3.4 Explaining the test score gap

The difference between the test score distributions in Finland and Germany may be due to several reasons. First, Finnish students may have a more favorable endowment in characteristics measured by the explanatory variables. Finnish students might for example have better

educated parents, who exert a positive influence on the test scores of their children, more resources at schools and a better institutional setting.

Second, the effects of the different characteristics on the test scores of students might differ between the two countries. In other words, the same characteristics might be less effective in producing reading literacy scores in Germany than in Finland. A greater return to family background characteristics in terms of test scores would imply a higher social differentiation of students. The educational achievement is then predetermined to a higher degree by the family of a student and students from lower social backgrounds find it harder to perform well at school. The effects of resources and institutions instead display the ability of schools to transform their endowment and their responsibility into improved student test scores.

Third, a part of the test score gap may be due to the difference in the residuals of the estimated regressions. Any unobserved factors that affect test scores, foremost innate ability of students and their motivation, constitute the residual effect. As the expected value of the residuals is zero, the residual effect is only important when one considers the test score gap at other points of the score distribution than the mean.

These three effects, referred to as the characteristics, the return and the residual effect can be quantified by decomposition methods. Two different methods are employed: the Oaxaca-Blinder (section 3.4.1) and the Juhn-Murphy-Pierce (section 3.4.2) decomposition.

3.4.1 Oaxaca-Blinder decomposition

This 'classical' decomposition technique has been developed by Blinder (1973) and Oaxaca (1973) and splits a gap into two parts. The first part is explained by differences in the characteristics, the second part by differences in the returns to those characteristics that have been estimated in the regressions. The Oaxaca-Blinder decomposition is performed at this point to give an overview over the average effects. Any differences in the effects along the test score distribution are examined in section 3.4.2.

The decomposition method used here differs slightly from the classical Oaxaca-Blinder decomposition and follows Lauer (2000). As the aim of the analysis is to explain the low test scores of German relative to Finnish students, the different effects that explain the gap in test scores are considered from the point of view of German students.

The total score gap between Finland and Germany at the mean is defined as:

$$(3.2) \quad \Delta T = \bar{T}_F - \bar{T}_G,$$

where the bars denote weighted averages and the subscripts F and G the countries Finland and Germany, respectively.

The total score gap can then be decomposed into a characteristics, a return and a characteristics-return effect, using the estimates of the weighted survey regressions presented in the previous section:

$$(3.3) \quad \Delta T = \sum_{i=1}^5 (\bar{X}_{iF} - \bar{X}_{iG}) \hat{\beta}_{iG} + \sum_{i=0}^5 \bar{X}_{iG} (\hat{\beta}_{iF} - \hat{\beta}_{iG}) + \sum_{i=1}^5 (\bar{X}_{iF} - \bar{X}_{iG}) (\hat{\beta}_{iF} - \hat{\beta}_{iG}),$$

where X comprises the five categories of explanatory variables G , B , R , I and S . The first component on the right hand side of equation (3.3) is the characteristics effect, which is the sum of the characteristics effects for all five categories of explanatory variables $i=1, \dots, 5$ included in equation (3.1). It measures by how much German students would score differently if they had the same characteristics as Finnish students, given their estimated returns to characteristics in terms of scores. The second component, the return effect, shows by how much German students would hypothetically score higher, if they experienced the same production process of schooling, i.e. the same transformation of inputs into educational achievement as Finnish students, given their own characteristics. It also includes the difference in the intercepts. Note that the difference between the estimated coefficients for the two countries is multiplied by German characteristics because the analysis considers the possible gain in test scores from a different production process for German students. In the original Oaxaca-Blinder decomposition, the Finnish characteristics would have been used as weights for the return effect. Then, the return effect cannot be interpreted from the point of view of German students. This modification in the decomposition leads to a third term explaining the gap in test scores. The characteristics-return effect is an interaction between the differences in returns and characteristics between the two countries. It expresses by how many test score points German students would score differently if Finnish students had not the advantage of being more endowed with those characteristics which yield higher returns in Finland than in Germany or less endowed in those characteristics that are transformed into higher test scores in Germany than in Finland.

The gap between the weighted average reading scores amounts to 54.29 points, as can be seen in Table 3.1, which presents the results of the decomposition. The analytical standard errors are computed according to Jann (2005). The difference in test scores is substantial since it is more than half of the international standard deviation of the PISA scores and is 44 percent higher than the effect of being in the ninth instead of eighth grade in Germany. The total characteristics effect is significantly negative, implying that German characteristics are actually more advantageous than the Finnish ones. The overall return effect of 58 points is significantly positive. The transformation of given inputs in Finnish schools results in higher student test scores than in Germany, which explains part of the score gap. The interaction effect amounts to 33 points.

A separation of the effects into the five groups of explanatory variables, grade level and age, student background, resources, institutions and school types, shows a more differentiated picture. The share of students in higher grades is larger in Germany than in Finland, leading

to the significantly negative characteristics effect of the grade category. While student background hardly differs between the samples of Finnish and German students, the difference in resources explains about four percent of the positive gap in test scores but is not significant. The percentage of high educated teachers is higher in Finland than in Germany. Instead, the negative effect for institutions implies more favorable characteristics for German students. In particular, the share of schools that are allowed to select their students is higher in Germany than in Finland. The negative characteristics effect for school types reflects that the gain in test scores from those students that attend schools with a positive school effect is higher than the negative impact of those students attending schools with a negative school effect relative to comprehensive schools in Germany.

The highly positive return effect is mainly driven by the difference in the intercepts and the returns to the resource variables. The return effect for student's age is positive, while it is negative for the two grade level variables. Since there are no students in grade ten in Finland, it is assumed that the return to being in grade ten compared to grade nine is zero in Finland. Excluding the tenth grade level variable would lead to a positive return effect for the grade category. The positive return effect indicates that resources are transformed more effectively into student test scores in Finland than in Germany. However, as the return effect for the resource variables is not significant, the results have to be interpreted cautiously. Moreover, the coefficient for instruction time is assumed to be zero in Finland because it could not be estimated. The transformation of personal and family characteristics is more beneficial for German than for Finnish students and more than offsets the positive effect for resources. However, this is due to the school type dummies that are included in the regression. The findings of a decomposition using only student background variables as explanatory variables are presented in section 3.4.2 as well. The highly positive difference of the intercepts between Finland and Germany indicates that a large part of the gap is due to unobserved factors.

For the decomposition, comprehensive schools are chosen as the reference category for school types. Hence it is assumed that Finnish schools are similar to German comprehensive schools. When the comprehensive Finnish schools are compared to a higher scoring German school type instead of the German comprehensive schools, the difference between the intercepts decreases. The choice of the reference school type only affects the three effects for the category school type and the difference between the intercepts. When no reference school type is specified and the production functions are estimated without intercepts, the return effect for schools turns more negative while the interaction effect for schools increases greatly. When only grade level and student background variables are used for the decomposition, the return effect for the constant term amounts to 19 test score points.

The significant interaction effect shows that Finnish students are more endowed with those characteristics for which they have higher returns than German students or vice versa. This is the case for all categories of variables except for resources.¹³

13. Similar to the return effect, it is also assumed for the calculation of the interaction effect that the return for being in tenth grade and for instruction time are zero in Finland.

Table 3.1
Decomposition for reading scores for all coefficients

	Sum	Grade	Student Background	Resources	Institutions	Schools	Cons
Total gap	54.29 (5.92)						
Characteristics effect	-36.39 (8.42)	-7.54 (1.26)	-1.32 (1.44)	1.94 (3.40)	-9.24 (4.04)	-20.23 (5.81)	
Return effect	57.81 (6.49)	1.06 (4.07)	-28.05 (6.24)	10.80 (29.90)	-7.76 (7.43)	-20.23 (5.81)	101.99 (31.89)
Interaction effect	32.87 (9.03)	10.25 (1.09)	4.73 (1.91)	-5.84 (5.47)	3.50 (4.96)	20.23 (5.81)	

Oaxaca-Blinder decomposition of test score gap between Finland and Germany. Germany is country of reference. Decomposition includes additional interaction effect (see equation (3.3)). Explanatory variables are divided into five groups. Standard errors in parentheses.

Table 3.1 considers all coefficients for the decomposition, even when the difference between coefficients in the two countries is not statistically significant at a reasonable level. Table A3.5 presents the decomposition results when those coefficients that do not differ at the ten percent-significance-level are restricted to be equal.¹⁴ Accordingly, only effects that significantly differ between countries are taken into account. These variables are: student's age, sex, parents' origin, parents' higher sec. education and all school types. For this adapted analysis, the effects in Table A3.5 differ only greatly for the resource and institutional variables, for which no coefficients differ significantly between the two countries. The sum of the effects hardly changes. The decomposition has also been conducted excluding the institutional variable on the selection of students by schools, which might be endogenous (cf. section 3.3.1). The results differ only slightly from those presented in Table 3.1 and the direction of the effects is identical.

When the average of the distribution is considered, differences in characteristics cannot explain the higher test scores of Finnish students. According to this decomposition, the inefficient transformation of the available resources and the difference in unobservable factors constitute the main reasons for the relatively low scores of German students, although the difference in the returns to the resource variables is not significant on average.

3.4.2 Juhn-Murphy-Pierce decomposition

Until now, only the mean of the distribution has been considered. However, as shown in section 3.2, the distribution of scores differs between the two countries. Therefore, the decomposition is performed along the entire score distribution as well.

14. The effects for the two countries have been estimated simultaneously using interaction terms to see if the coefficients for the countries differ significantly. The interaction terms that are not significant have then been dropped. Reducing the significance level to five percent leads to a further reduction of coefficients.

The following decomposition technique is based on a decomposition that was first employed by Juhn et al. (1993) for a decomposition of change across time. The method has the distinct advantage of considering not only the mean for the decomposition but the whole distribution. Moreover, it deals explicitly with the residuals from the estimations of the production function, which are equal to zero at the mean but not at specific points of the distribution. The Juhn-Murphy-Pierce decomposition has been applied to cross-sectional data like the PISA data before (e.g. Blau and Kahn, 1996; 2005). Following the decomposition approach in Lauer (2000) as depicted in equation (3.3), the decomposition allows one to decompose the score gap into a characteristics, return, characteristics-return and residual effect. The decomposition that is used here hence differs from the original Juhn-Murphy-Pierce decomposition in two points: it decomposes the difference between two cross-sectional distributions instead of changes across time and it splits the difference in four instead of only three different effects, the characteristics-return or interaction effect being added.

The residual ε_i of country y can be thought of consisting of two components: the percentile of an individual i in the residual distribution θ_i , and the distribution function of the residuals, H_i . The inverse cumulative residual distribution function then gives us:

$$(3.4) \quad \varepsilon_{iy} = H_y^{-1}(\theta_{iy} | X_{iy}),$$

where X comprises the five sets of explanatory variables G , B , R , I and S of country y . Using the estimates from weighted survey regressions of equation (3.1), the actual and two hypothetical test score distributions for German students can be constructed:

$$(3.5) \quad GER_i = X_{iG} \hat{\beta}_G + H_G^{-1}(\theta_{iG} | X_{iG})$$

$$(3.6) \quad GER(1)_i = X_{iG} \hat{\beta}_F + H_F^{-1}(\theta_{iG} | X_{iG})$$

$$(3.7) \quad GER(2)_i = X_{iG} \hat{\beta}_G + H_F^{-1}(\theta_{iG} | X_{iG}).$$

The first hypothetical distribution GER(1) shows the scores that German students would attain if they experienced the Finnish returns to their characteristics and the corresponding residuals from the Finnish residual distribution. The residuals are the outcome of the inverse of the Finnish cumulative residual distribution function H_F for the percentile of individual i in the German residual distribution, θ_{iG} , given the characteristics X_{iG} . Equation (3.7) presents the second hypothetical distribution GER(2), which assumes that the characteristics of German students are transformed into test scores by the German returns, but that the residual distribution is the same as for Finnish students. The two hypothetical Finnish distributions are created likewise.

Following the decomposition as described in equation (3.3), the characteristics effect is the difference between the test score distributions for FIN(1) and GER. The return effect equals the difference between the two hypothetical test score distributions GER(1) and GER(2). The third effect is due to the different distributions of residuals in the two countries and

can be calculated by subtracting GER from GER(2). The interaction effect can then be constructed as $(\text{FIN}-\text{FIN}(1))-(\text{GER}(1)-\text{GER})$. Adding up all four effects leads to the total gap $(\text{FIN}-\text{GER})$ that shall be explained here. The resulting score distributions are discussed in the following section.

The hypothetical test score distributions

In order to show the different effects graphically, Figure A3.1 in the appendix displays the real and hypothetical test score distributions. The differences between the reading score distributions as estimated by the kernel density function are only due to one of the following effects. The first effect in the upper left hand graph of Figure A3.1 is the characteristics effect. The graph shows the hypothetical Finnish distribution FIN(1) and the actual German distribution GER. The former one displays how the distribution would look when Finnish students with their own characteristics experienced the German returns to these characteristics and the German residuals given their position in the Finnish residual distribution. The difference that remains between the two distributions is only due to differences in characteristics between the two countries, given the German returns. The mode of the hypothetical Finnish distribution is positioned to the left of the German distribution, which has a higher spread and is slightly skewed to the left. This implies that the German characteristics are actually more favorable (according to the estimation results) and heterogeneous compared to the Finnish characteristics. The characteristics effect thus implies higher average test scores for the German than for the Finnish characteristics. However, in the lower part of the distribution, the size of the effect decreases and implies higher scores for Finnish than for German characteristics. This is consistent with the negative slope of the total score gap over the distribution.

The return effect is shown in the upper right hand side of Figure A3.1. The hypothetical distribution GER(1) shows the predicted scores for students with German characteristics, Finnish returns and the Finnish residual distribution. Distribution GER(2) displays how students with German characteristics and German returns but with Finnish residuals would score. The difference between the distributions is only due to differences in the returns to characteristics in the two countries, given the German characteristics. The educational production process in Finland clearly leads to higher test scores of students, especially for the lower part of the distribution. The return effect can hence partly explain why German students are scoring lower than Finnish students.

The residual effect is depicted in the lower left hand graph of Figure A3.1, where the distributions GER(2) and GER are compared. The Finnish residuals in GER(2) lead to a distribution with a larger variation than the German residuals. This is consistent with the earlier results on the determinants of test scores, which showed that the observable factors can explain a higher share of the variation in test scores in Germany than in Finland. Consequently, unobservable factors like innate ability of students have a greater impact in Finland, which is implied by the residual effect. The last effect in Figure A3.1 is the interaction effect in the lower right hand side, which is positive. It shows that German students would score higher if Finnish students had not the advantage of being more endowed with those characteristics which yield

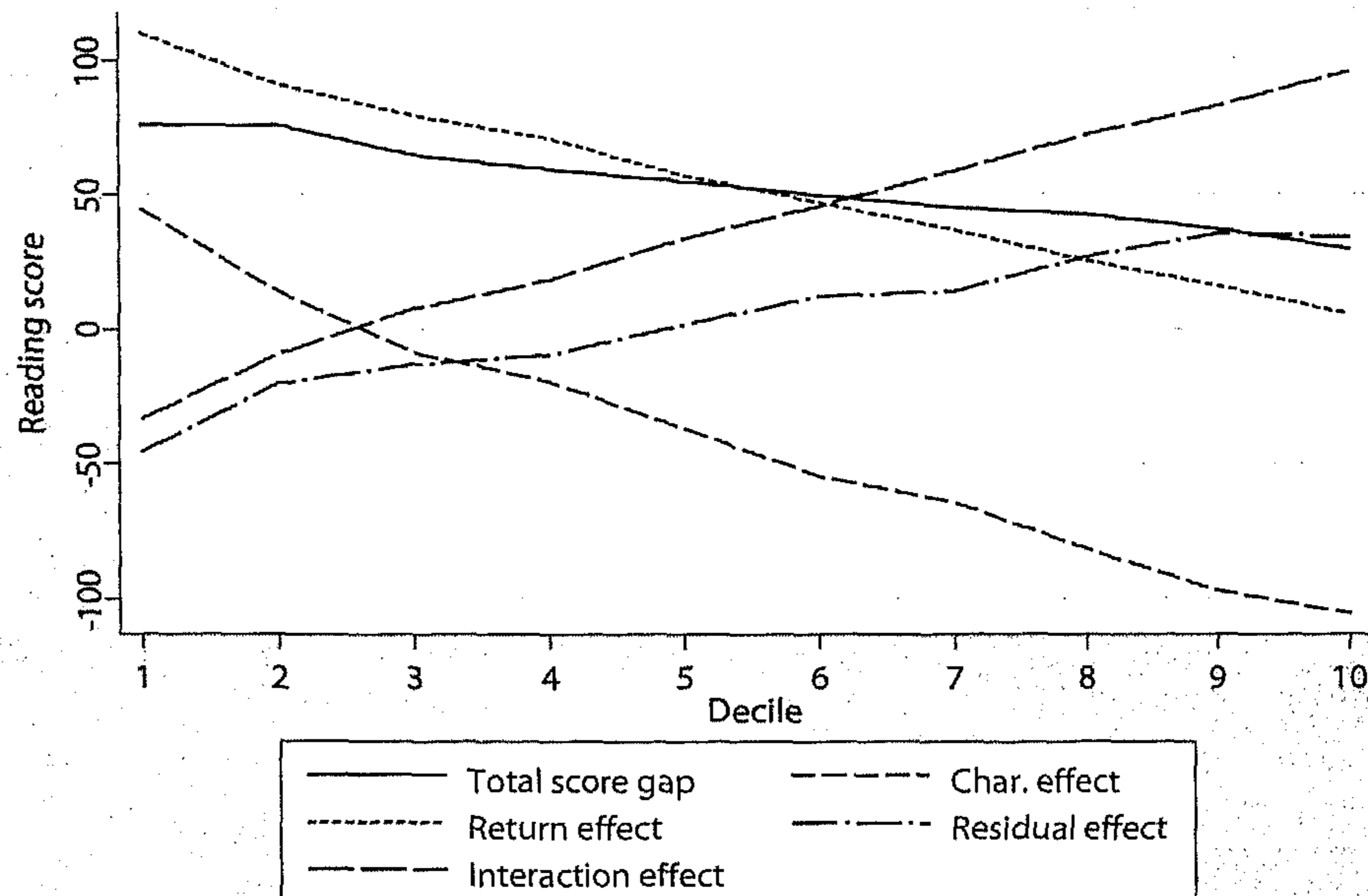
higher returns in Finland than in Germany or less endowed in those characteristics that are transformed into higher test scores in Germany than in Finland.

The hypothetical distributions showed the predicted test scores for German and Finnish students had they experienced the returns and residuals of another educational system. The following section pays a closer look to the different effects and the contribution of the different groups of variables to the effects.

The effects and their components

The four effects, which make up the total score gap between the countries, can be broken down further and linked directly to the five groups of variables that determine student test scores. Since the average effects have already been discussed in the previous section, the focus is now on the course of the effects along the test score distribution. First, the four effects are shown over the deciles of the test score distribution. Figure 3.3 displays the score gap and the total effects.

Figure 3.3
Total effects of decomposition



The characteristics effect can explain part of the test score gap, but only for the lowest three deciles of the test score distribution, where it is positive. It decreases steadily, implying that the characteristics are deteriorating comparatively more in Germany than in Finland when going down the score distribution. Hence, the measured characteristics are more strongly linked with students' test scores in Germany than in Finland. The return effect decreases as well over the distribution, but is always positive. The problem of converting the given endowments in Germany into high test scores of students is hence greatest for the lowest scoring students. The residual and interaction effect run almost parallel and oppose the other effects. They increase over the whole distribution and are positive from about the fifth and third decile upwards, respectively. The increase in the residual effect is caused by a steeper rise in the Finnish residuals that are first lower and then higher than the German residuals. This implies that unobservable factors explain more of the variation in test scores in Finland, which can also be seen in the lower left hand part of Figure A3.1, since students at the bottom of the distribution have lower residuals and students at the top have higher residuals than the corresponding German students. The interaction effect increases because the German characteristics deteriorate more when moving down the score distribution than the Finnish characteristics. When a higher scoring German school type than the comprehensive school is chosen as the reference category for school types, the total characteristics effect shifts upward while the interaction shifts downwards. Both shifts are only due to changes in the category school types. The confidence intervals of the effects based on bootstrapping with 5000 replications using schools as sampling clusters are presented in Figure A3.2. The 95 percent confidence intervals show that all effects change significantly along the distribution and differ significantly from zero at some point of the distribution.

Now we turn to the composition of the effects. Figure 3.4 displays the five components of the characteristics effect. While the positive effect of resources and the negative effect of institutions does not vary greatly over the distribution, the characteristics effects of grade level and student background decrease and turn from positive for the lowest three deciles to negative.¹⁵ Low scoring students in Germany have hence less and high scoring students more favorable characteristics than Finnish students at the same decile of the score distribution. The characteristics effect for the variables describing the type of school is also positive for the lower part of the distribution but decreases very strongly up to -70 for the highest decile. The streaming of the schooling system is hence associated with a greater inequality between students in Germany than in Finland because it introduces an additional source of test score variation. It is unknown to which degree the effects of school type are due to sorting or due to genuine school type effects, though (cf. section 3.3.1). The level of the characteristics effect for school types depends on the choice of the reference school type. It increases when a higher scoring German school type is chosen as reference category instead. Figure A3.3 presents the confidence intervals for the components of the characteristics effects. Except for the category of resources, the characteristics effect is significantly different from zero for at least one decile.

15. The student background variables that change the most along the score distribution are parents' education and books at home and less the personal characteristics like student's sex.

Figure 3.4
The components of the characteristics effect

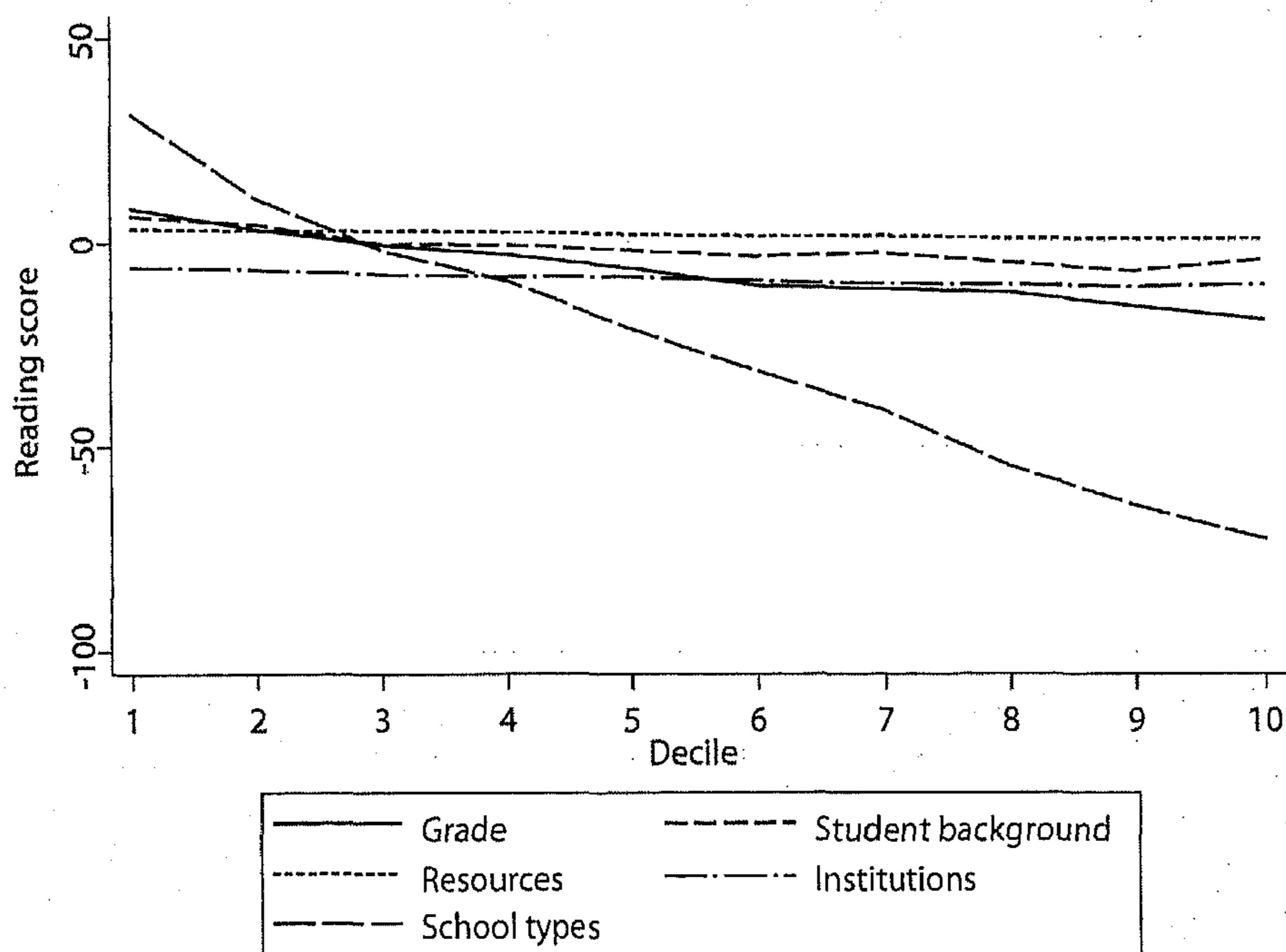
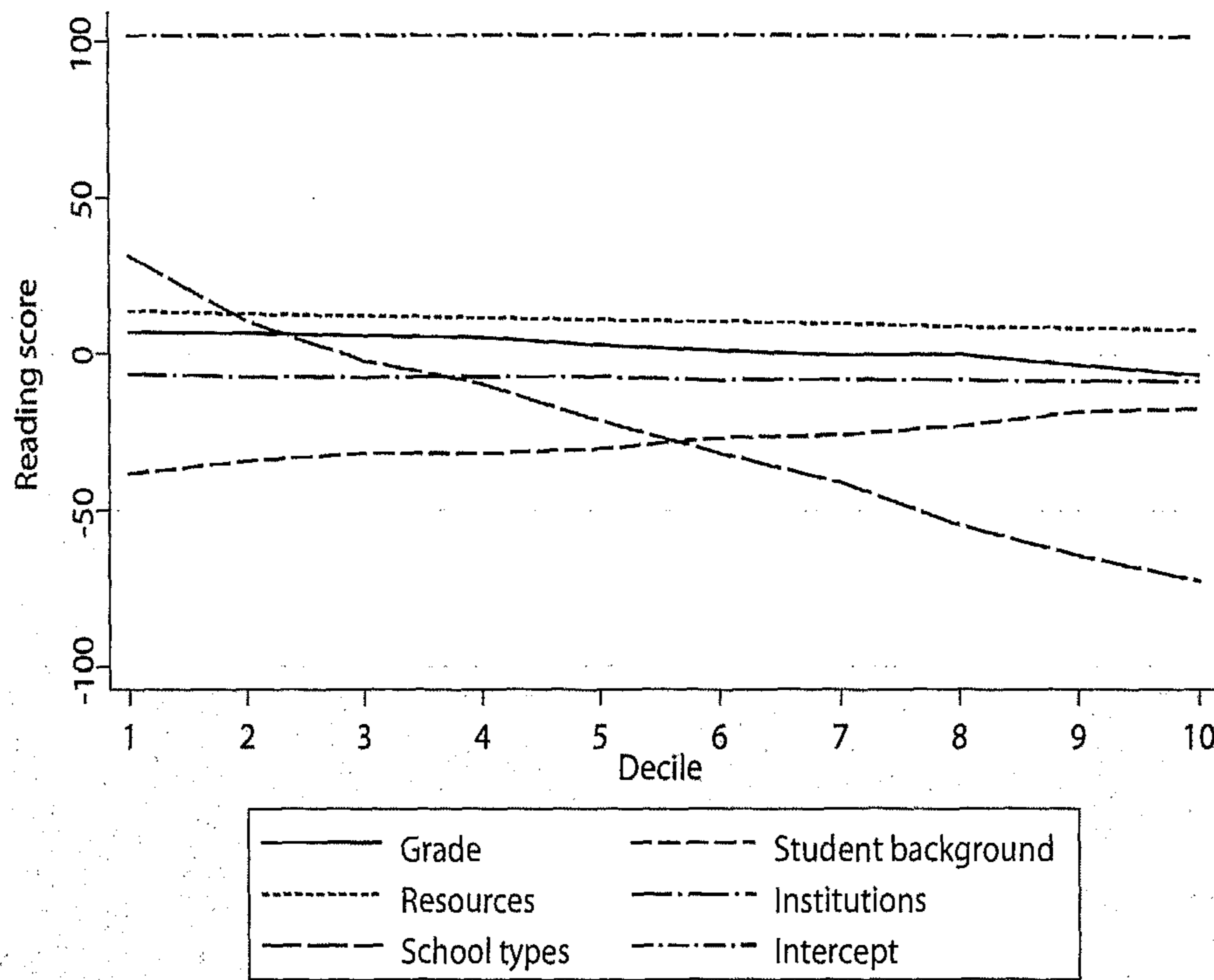


Figure 3.5 shows the return effect separately for each group of variables. The negative return effect for student background increases over the distribution. However, it remains unclear how much of the effect of student background on test scores is hidden in the school type variables in Germany (see section 3.3.1). The return effect for resources is slightly positive along the whole score distribution and can hence explain the score gap partly. Institutional variables contribute negatively to the return effect at all deciles. The return effect for school type variables decreases along the score distribution and is negative from the third decile onwards. Since Finnish students are all in comprehensive schools, the effect reflects only that students in the lowest three deciles are more likely to attend low secondary schools (Hauptschule) in Germany, which have a negative effect compared to the reference group of comprehensive schools, while higher scoring students are more likely to be in medium (Realschule) or higher (Gymnasium) secondary schools. The difference in the intercept is highly positive and constant along the distribution, implying that the level of test scores is higher in Finland than in Germany due to unobserved factors and the choice of the reference school type. When a higher scoring German school type is chosen as reference category instead, the return effect for school types increases while the difference in the intercepts decreases. Figure A3.4 shows that the return effect does not differ significantly from zero for the categories grade, resources

and institutions at any decile of the distribution. The confidence interval for the difference of the intercepts, which is not shown in Figure A3.4, is very wide but always significantly positive.

Figure 3.5
The components of the return effect



The decomposition was also performed for the grade and student background variables only in order to test the robustness of the results when comparing a schooling system where streaming is used (Germany) to a single school system (Finland). The findings are in line with the overall results presented here but oppose the results for the group of student background variables. There is a negative characteristics and a positive return effect. When looking at the individual grade and student background variables, the characteristics in Germany are less favorable compared to Finland concerning student's age and parents' origin but more favorable concerning the grade level, parents' education and number of books at home. The return effect implies that there are more favorable returns in Germany concerning the sex of students and parents' origin but unfavorable returns to student's age, parents' education and the number of books at home than in Finland. This shows once again that the differ-

ence in test scores between students with a low and high social background is much larger in Germany than in Finland.

In order to control for a possible bias of the results through a selection of students into grades (see sections 2.2.1 and 3.2), the decomposition analysis is repeated drawing on a sample of students in grade nine only. The majority of 15-year-old students in both countries are in grade nine. Therefore, students in grade nine should not be affected by selection through promotion rules. The results of the decomposition for this reduced sample are very similar to the results presented above, both for the Oaxaca-Blinder and Juhn-Murphy-Pierce decomposition. The overall effects as well as the effects for the various categories of explanatory variables are identical in their direction and differ only slightly in their magnitude.¹⁶ The decomposition has also been conducted excluding the institutional variable on the selection of students by schools, which might be endogenous (cf. section 3.3.1). The results differ only slightly from those presented in the figures and the direction of the effects does not change.

A possible further step in the decomposition analysis would be to consider the estimated coefficients along the conditional distribution, not only at the mean of the distribution (cf. Garcia et al., 2001). However, most coefficients estimated by quantile regressions do not differ significantly from OLS coefficients. Only the coefficient for student's age increases while the coefficient for student's sex decreases slightly but significantly over the conditional test score distribution for Germany. Overall, a decomposition using quantile regressions apparently does not add further insights.

3.5 Conclusion

The decomposition analysis showed that the gap in test scores between Finnish and German students cannot be explained by differences in student background and school characteristics, except for the bottom of the score distribution. German students have on average a more favorable background but the impact of having lower background characteristics is much stronger for German than for Finnish students. This holds in particular for the social background of students but not for student characteristics such as sex and the grade level of students. The measured institutions are more favorable in Germany while Finland is endowed with slightly more resources. Resources are used more effectively in Finland than in Germany but the difference is not significant. A large part of the overall score gap between the countries is due to unobservable factors. The results might imply that streaming in Germany affects students in lower school types negatively and leads to greater inequality of educational achievement. It remains unclear, however, if this can be attributed to the effect of school types per se or student background and innate ability that determine the allocation process of students into school types. Overall, the variation in test scores can be explained much better by observable characteristics in Germany than in Finland.

¹⁶ The results of the robustness checks are available from the author upon request.

In order to improve the performance of students in Germany as measured by test scores, especially the educational achievement of students in the lower part of the test score distribution has to be promoted. These students suffer from a highly disadvantaged student background, whose negative impact upon test scores might be intensified by the early streaming in the German schooling system at the age of ten. They have little possibility to compensate for their background before they are divided into different school types. There is no evidence for a significant beneficial effect of higher teacher-student ratios or other resources.

The large difference in the degree of equality in educational opportunity between countries that has been observed in this analysis and the role of institutions are analyzed in a cross-country analysis in the following chapter.

4

Educational opportunities and the role of schooling institutions

This chapter analyzes the effect of schooling institutions on educational opportunities. It complements the empirical research of this thesis by investigating the impact of the differences across schooling systems at the country level.¹ The theoretical model on the relationship between schooling institutions and the degree of equality of educational opportunity that was developed in section 2.1 serves as a basis for the analysis. The current chapter empirically investigates this relationship and aims at validating the hypotheses on the institutional effects from the theoretical model.

4.1 Introduction

Based on the review of the literature and the previous theoretical and empirical analyses, this chapter pursues two main objectives. First, it intends to show what effect student background has on student performance as measured by test scores at two stages in a student's life. This describes the degree of equality of educational opportunity that students face at different stages within a particular schooling system. Besides social background also the attitude of parents towards their child's education and student characteristics are considered in order to describe the various dimensions of educational opportunity. Second, the chapter aims at exploring whether cross-country differences in educational opportunities are related to essential features of educational systems.² This step of the analysis provides empirical evidence on the link between institutional settings and educational opportunities, and seeks to elaborate on better frameworks for equal opportunities. The analysis builds on internationally comparable micro data from the two studies PIRLS (Progress in International Reading Literacy Study) and PISA (Programme for International Student Assessment) on student performance in 14 countries, which have been introduced in section 2.2.1. The schooling systems are analyzed at grade four (PIRLS) and grade nine (PISA), two important points in a child's development

-
1. I would like to thank John Bishop, Lex Borghans, Bernd Fitzenberger, Kevin Lang, Charlotte Lauer, François Laisney, Alexander Spermann, Ludger Woessmann and participants at the ESPE meeting in Paris, SOLE/EALE meeting in San Francisco and ROA seminar in Maastricht for helpful comments on this chapter.
 2. Educational opportunities are defined as the effect of student background variables on student performance measured by students' test scores in the PIRLS and PISA studies.

and in the schooling system.³ The first point is associated with the end of primary education in many countries while the second point often constitutes the end of compulsory education. The impact of the following schooling institutions is analyzed: the number of school types / use of streaming in school systems, annual instruction time, share of students in private schools and school autonomy. These institutions have been chosen because they are likely to affect educational opportunities of students rather than student performance and are at the center of the debate on how to improve the equality of educational opportunity in many countries. Moreover, the institutions vary between the countries considered here but hardly within countries and are hence well-suited for this cross-country analysis and are also analyzed in the theoretical part (see section 2.1).

The role of institutions in determining the quality of schooling has been investigated by several recent papers (see section 1.1.2 for a discussion of the literature), which were also based on international studies on student performance (e.g. Vandenberghe and Robin, 2004; Woessmann, 2003). The approach followed in most studies is to determine the effect of institutional settings on student performance by estimating educational production functions (e.g. Fuchs and Woessmann, 2006; Schuetz et al., 2005) or applying a matching approach (e.g. Vandenberghe and Robin, 2004). However, in a cross-sectional international comparison one cannot perfectly control for cultural and societal differences between countries. Therefore, only institutional features that vary within countries can be reasonably analyzed by these approaches (cf. Ammermueller, 2006). The effects of institutional settings like the structure of schooling systems, the length of the school year or other features that apply to all students within a country are likely to depend on unobservable factors like cultural background because they developed over a long time span within the cultural framework of the country. Hence, the effect of institutions – be it on the level of student performance or on the degree of educational opportunity – cannot be consistently estimated in such cross-country analyses. One approach to overcome these difficulties is to look at difference-in-differences evidence from two cross-sectional studies on student performance.⁴ Hanushek and Woessmann (2006) follow this approach and assess the impact of streaming in secondary schools on overall inequality. They find that test score inequality increases between primary and secondary education in countries with early streaming schooling systems compared to late streaming systems. However, they focus only on one institutional factor and one measure of inequality and do not exploit the micro-level data. Thereby, the effects on the different dimensions of educational opportunities cannot be examined and possible other sources of inequality are ignored by their approach, leading to a potential omitted variable bias. The study by Waldinger (2006) is similar to the approach followed here but considers only the timing of streaming and no further institutional features. Using a more restrictive specification and not accounting for the degree of streaming, he finds positive but no significant correlations between streaming and the importance of family background on test scores. Schuetz et al. (2005) is the only study that investigates the impact of several institutional features of a schooling system as well but controls for country-specific effects to a lesser degree compared to this study. Meghir and Palme (2005) make use of a natural experiment of a

3. 'Grade nine' refers to the PISA data and includes all 15-year-old students in grades eight, nine and ten.

4. There is no cross-country panel data available on student performance.

reform in Swedish school districts to assess the impact of a shift from a streamed to a unified secondary school system on educational attainment and earnings. They find that educational opportunity increased as a result of the reform. The reform included as well further changes like an extension of compulsory schooling, which makes it difficult to contribute the findings to a specific change in the institutional setting. The same critique applies to the evaluation of the Finnish school reform that led to an increase in intergenerational income mobility (Pekkarinen et al., 2006).

The estimation strategy employed here aims at a consistent estimation of the link between institutions and educational opportunities using both PIRLS and PISA data. A second approach is applied in addition to a cross-sectional approach in order to explain changes in educational opportunities over time (measured by student age) by changes in institutional features. Thereby, country-specific factors besides the schooling system can be largely controlled for, assuming they are identical for students of age ten and fifteen. Attention is also paid to the comparability of cross-sectional student performance studies.

The estimation of the effects of student background on student test scores shows that the absolute effect of gender, amount of books at home, school location and parents' attitude increases between the end of primary and lower secondary education in most countries, while the effect of parental education and parents' origin seems to decrease. Most variables reflecting the behavior of students and parents are gaining in importance while most family characteristics are becoming less decisive for determining student performance. Therefore, educational opportunities seem to increase for individuals with a positive attitude towards education. Moreover, the results show that educational opportunities are not one-dimensional. There is a considerable amount of variation in the changes of educational opportunities over a student's schooling career between countries. It remains unclear what part of the changes in educational opportunities is due to differences between the studies PIRLS and PISA and what changes really take place, though.

It can be shown that schooling institutions that are determined by school policy are linked to educational opportunities of students. Most of the results of the theoretical model are supported by the empirical evidence. The social origin of students, measured by the number of books at home and parental education, has a larger effect on student test scores rather in countries with a differentiated secondary schooling system with several school types and in countries that experience an increase in the size of the private school sector between primary and secondary education. This supports the hypothesis that streaming benefits the performance of students from a higher social background. The time students spend in schools seems to limit the effect of social origin upon student test scores while school autonomy is positively linked to the impact of parental influence.

The chapter is structured as follows. Section two introduces the two data sets and compares them. The estimation strategy is outlined in section three. Section four discusses the results of the estimations of the educational production functions while section five presents the

evidence on the role of the schooling systems. The final section summarizes the results and concludes.

4.2 Data

The data from two international studies on student performance in reading literacy are taken for the analysis of 14 countries. The Programme for International Student Assessment (PISA) tested 15-year-old students while the Progress in International Reading Literacy Study (PIRLS) refers to the performance of students in grade four (ages nine and ten). Both datasets have been introduced in section 2.2.1. The countries that are included in both studies and provide the necessary background information are Canada (CAN), the Czech Republic (CZE), England (ENG), France (FRA), Germany (DEU), Greece (GRC), Hungary (HUN), Iceland (ISL), Italy (ITA), Latvia (LVA), New Zealand (NZL), Norway (NOR), Russia (RUS) and Sweden (SWE). The analysis is conducted both including and excluding Canada and Great Britain, for which the sampled regions differ between PIRLS and PISA.⁵ Information on the participation in PIRLS and PISA on school and student level for both studies is given in Table A4.1. Tables A4.2 and A4.3 present the means and standard deviations for the data. The following sections describe both studies and discuss the handling of missing values and the comparability of the data in addition.

Figure A4.1 in the appendix displays the mean PIRLS scores and their standard deviations for all 14 countries in a scatter plot. The negative trend line implies that countries with higher mean test scores tend to have a lower spread of scores but the relationship is not significant. The high standard deviations in New Zealand and England are striking. The PISA scores used for the analysis are plausible values as in the PIRLS data and are standardized in the same way. However, the sample of countries differs for the two studies. Figure A4.2 shows that no relationship between mean test scores and the standard deviations seems to exist across countries in PISA. Germany is an outlier with an extremely high spread of scores.

4.2.1 Imputation of missing values

Missing values for student background variables are a problem of the data from both studies (see also section 2.2.2). Test scores are reported for all students but some students did not complete the tests. Table A4.4 in the appendix presents the percentage of missing values for all variables and countries. Commonly, the whole observation (student) is dropped from the regression whenever the value of any explanatory variable is missing. In ENG, 65 percent of the students in PIRLS would have been dropped for example. Apart from losing valuable information, dropping students with incomplete answers to the questionnaires might lead to a sample selection bias if the values are not missing randomly conditional on student

5. In PIRLS, Canada is represented only by the provinces of Ontario and Quebec and only England is sampled while in PISA, the whole of Canada and Great Britain are sampled. The United States participated in both studies but provides no information on parents in PIRLS. The U. S. is included in the graphs comparing the two studies but not in the later analysis.

performance. Indeed, given that attentive students are more likely to both complete the questionnaire and to answer the test questions, low performing students have a higher probability of being dropped. Thus, dropping observations with missing values leads to an upward bias in the test scores, which can be seen for all countries but Canada in Table A4.3 in comparison to Table A4.2.

The approach chosen here to overcome the problem of missing data is to predict missing values on the basis of regressions on those background variables like age, sex and the grade a student attends that are available for all students. Linear models are used for continuous variables and probit and ordered probit models for qualitative variables.⁶ Students who did not answer these elementary background questions or did not complete the tests have been excluded from the regressions. Observations with a share of missing values above 50 percent have been excluded as well.

The prediction of missing values on the basis of regression results is clearly no impeccable solution. The variation of the variables decreases, as can be seen in the lower standard deviations of the variables including the imputed values as compared to the original data (see Tables A4.2 and A4.3). However, the imputed values vary greatly as well and the information of the non-imputed values of the observation is kept. Several other empirical studies that apply this method to test score data show that the imputation has no effect on the qualitative results (e.g. Fuchs and Woessmann, 2006; Woessmann, 2004).

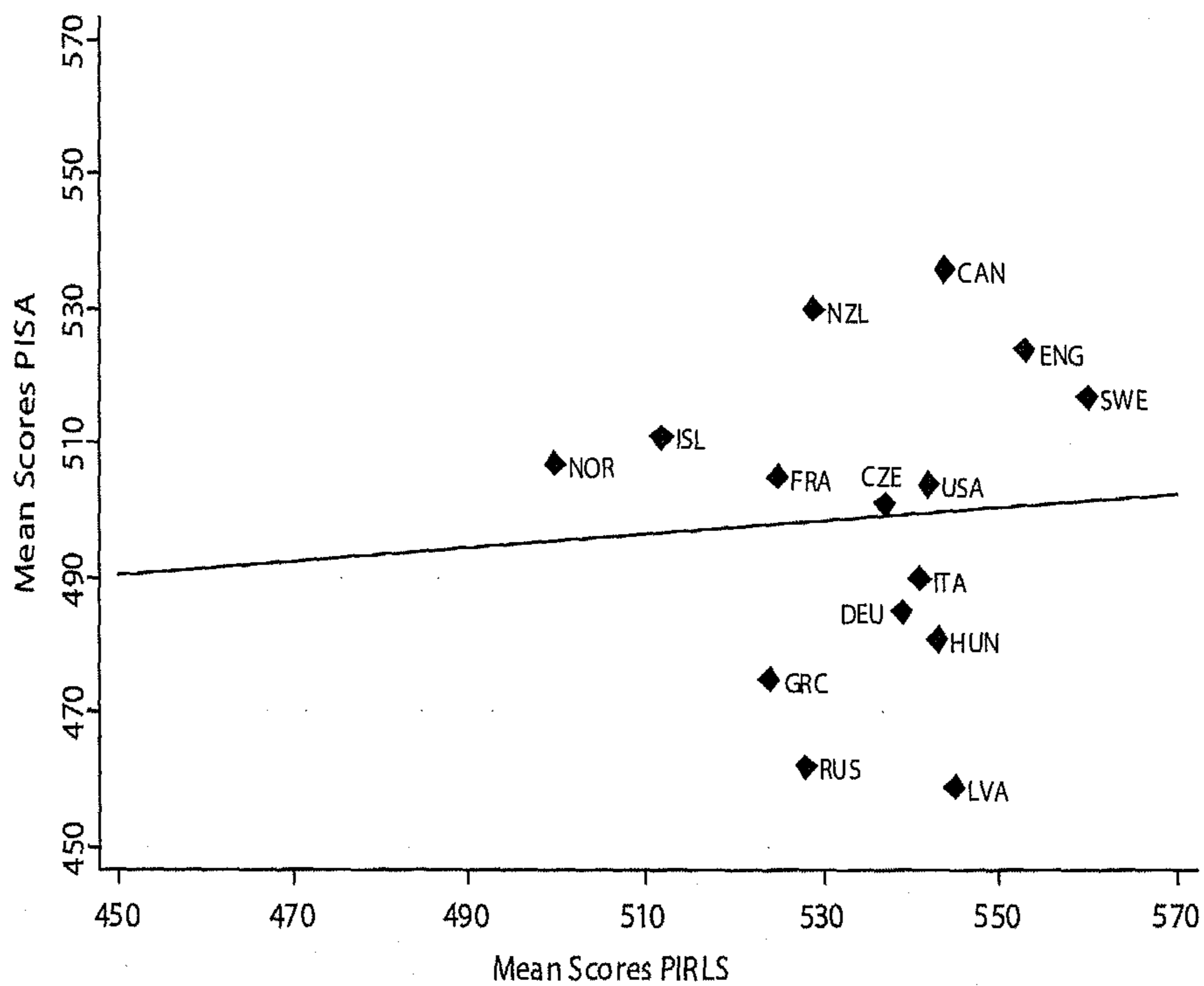
4.2.2 Comparability of the studies

The two studies have not been designed to be comparable to each other and they differ partially. However, both studies use the same concept, to test students' understanding rather than knowledge and they depend little on national curricula, making the studies internationally comparable. Both have an international mean of 500 and a standard deviation of 100. As the samples of countries differ, the country means of the scores are not directly comparable. For the countries considered in this study, the average PIRLS scores are mostly higher than the average PISA scores, which can be explained by low scoring non-OECD countries that are sampled in PIRLS but not in PISA. The regression line in Figure 4.1 shows that there is a slight positive relationship, which is not significant, between the PIRLS and PISA mean test scores on the country-level. Rescaling the scores such that the overall mean score and standard deviation are identical for PIRLS and PISA in the sample of countries used in the analysis assumes that there is no change in aggregate scores between grade four and grade nine. Since the correct scaling is unknown and any rescaling is arbitrary, the results presented in the following are based on the original scores but the whole analysis is repeated using rescaled scores in order to test the robustness of the findings (see section 4.4.5). The rescaled scores have been computed such that the overall mean of the scores equals 500 and the standard deviation 100 for the sample of 14 countries for both PIRLS and PISA. The description of the imputation methods for the plausible values of the test scores in PIRLS and PISA

6. For a more detailed discussion of the imputation method, see Woessmann (2003). The imputation method has already been used in Chapter 3 and is similar to the STATA `impute` command.

in the technical reports indicates that the methods are similar. Since the analysis refers to a rather homogeneous group of mostly OECD countries, the score distributions are affected by the imputation method of the test scores in a comparable way (cf. Brown et al., 2005). Still, it has to be assumed that possible minor differences do not bias the results. Hanushek and Woessmann (2006) show that the results are largely independent of the choice of student performance studies that are used in their difference-in-differences analysis.

Figure 4.1
Mean test scores in PIRLS and PISA



The relevant variables like parents' education and number of books at home have been transformed in the data in order to be comparable. All student background variables have been transformed into dummies except for student's age. The effects of the student background variables on test scores should therefore not only be comparable across countries but also across studies, assuming that test scores are comparable across studies. Table A4.5 presents the correlation coefficients between the country means from the two studies for variables which should have similar mean values, although the sample of students and their age differ. For the origin of parents, the number of books at home and the school location the correlation coefficients are reasonably high both for the mean values computed with and without the imputed values. The high coefficients for books at home indicate that this is a trustworthy

measure of home resources. The coefficients are quite low instead for the language spoken at home and some categories of parents' education, especially for the country means including imputed values. This may be due to the fact that different categories are used in the studies to report education and that the share of missing values for education is relatively high (see Table A4.4). In PIRLS, the share of secondary schooling is relatively high compared to PISA, in which the share of university education is relatively high for most countries. Differences in the definition of student background variables should have no effect on the estimates of interest of the second estimation approach unless they are systematically related to schooling institutions.

The studies have been conducted in 2000 (PISA) and 2001 (PIRLS). The cohorts of students that are tested are hence about six years apart. The lag of one year between the two tests should not affect the comparability because schooling institutions do not change strongly within the span of one year.

4.3 The effect of student background on student performance

This section shows how the impact of student background on test scores is estimated, which serves as a description of the degree of equality of educational opportunities. The results for both PIRLS and PISA are presented and compared to each other.

4.3.1 Estimation outline

Based on the theoretical model in section 2.1, the considerations on the empirical estimation and the data limitations, an educational production function is set up to measure the impact of student background on student test scores. In comparison to equation (3.1), the following equation that is estimated separately for each country and each study only includes variables describing student's background, i.e. school variables are excluded. Suppressing subscripts that refer to PIRLS/PISA and the country, the educational production function is written as:

$$(4.1) \quad T_{is} = \beta_0 + B_{is}\beta_1 + D_{is}\beta_2 + \eta_s + \varepsilon_{is}$$

where

T_{is} : the reading test score of student i in school s .

B_{is} : student background variables whose impact on test scores is potentially affected by institutions (cf. sections 2.1.2 and 4.4.2):

- number of books at home (1 if more than 200/250 books, 0 otherwise)⁷
- student's parents' education (1 if either parent finished university, 0 otherwise)
- student's parents' attitude (low and high) towards education (assessed by teacher/principal):

7. In the PISA student questionnaire, the categories of books at home differ slightly from the PIRLS student questionnaire: ten and fewer books, 11-50 books, 51-100 books, 101-250 books and more than 250 books at home.

- 1 if low, 0 otherwise
- 1 if medium, 0 otherwise (reference)
- 1 if high, 0 otherwise.

D_{is} : further student background variables:

- student's gender (female = 1, male = 0)
- student's age (in months)
- student's parents' origin (1 if at least one parent was born abroad, 0 otherwise)
- students' parents' education (the highest level achieved by either parent). For Canada no information is available on upper secondary education in PISA.
 - 1 if no finished upper secondary education, 0 otherwise (reference)
 - 1 if finished upper secondary education, 0 otherwise
 - 1 if finished post-secondary education, 0 otherwise. No distinction is made between upper and post-secondary education for Norway in PIRLS.
- language spoken at home (0 if home country's, 1 otherwise). No information available for Hungary in PISA.
- number of books at home (see footnote 36):
 - 1 if $\leq 25/50$, 0 otherwise (reference)
 - 1 if $26/51 - 100$, 0 otherwise
 - 1 if $101 - 200/250$, 0 otherwise
- school's location (size of community where school is located). For Canada and Iceland no data are available for school's location in PISA.
 - 1 if rural ($< 3,000$ inhabitants), 0 otherwise
 - 1 if town ($3,000 - 100,000$ inhabitants), 0 otherwise (reference). No distinction made between "rural" and "town" for Russia in PIRLS.
 - 1 if city ($> 100,000$ inhabitants), 0 otherwise
- student's grade (PISA only):
 - 1 if grade = 8, 0 otherwise
 - 1 if grade = 9, 0 otherwise (reference)
 - 1 if grade = 10, 0 otherwise.

ε_{is} is an error term at the student level. It is assumed independently distributed with constant variance for each country. η_s is an error term at the school level.

The vectors β_0 , β_1 and β_2 are to be estimated. The vector β_1 serves as a measure of educational opportunity (cf. Blanden et al, 2005; Woessmann, 2004). For the number of books at home and parents' education, only the highest categories are included in B . The reason is that an impact of institutions as laid down in the theoretical model should be more pronounced when considering the difference between the highest and lowest category of student background variables. By including only variables on student background and none on schools, β_1 estimates the total effect of student background on test scores. Any indirect effects of background on previous performance and school choice are included (also see the discussion of student background effects in section 3.3.1). Since school factors are omitted in this reduced form regression, β_1 should rather be interpreted as a correlation and not a causal effect. This correlation between test scores and student background is used as the measure for educational opportunities because the interest from a policy point of view is on differences in

performance between students from different social backgrounds. The difference in performance might be due to differences in e.g. innate ability, home resources, peers and schools. This measure of educational opportunities relates to the theoretical model in section 2.1 and is the common measure used in the related literature (Brunello and Checchi, 2006; Schuetz et al., 2005; Waldinger, 2006). Alternatively, equation (4.1) can be estimated by a school-fixed effects model, controlling for all differences between schools. Then β_1 estimates the effect of student background conditional on the school that students attend and thereby neglects the impact of school selection. The analysis refers to reading literacy because it is the only proficiency that is tested in both PIRLS and PISA.

The student background variables are unlikely to change much over time and are hence a good proxy also for prior inputs of student background. In schooling systems in which the enrolment age is not strictly regulated but depends on the characteristics of students or parents, the grade level dummies in equation (4.1) may be endogenous. Therefore, estimations using only one grade level are performed as well.

4.3.2 Estimation results

This section presents the estimates of student background effects on student test scores, including both the student background variables used to measure educational opportunity and further student background variables. Equation (4.1) has been estimated separately for both studies and each country using CRLR with schools as clusters. First, the results from PIRLS for fourth-graders, then from PISA for ninth-graders and finally the differences between the results are presented.⁸ It should be noted that, due to missing variables, restrictions have been imposed on coefficients in some cases. The coefficients of 'Upper Secondary' and 'Post Secondary' in the case of Norway in the PIRLS regression, for example, have been set equal to each other.

Effects at the end of primary education

Table 4.1 displays the estimates for the effect of student background variables on reading literacy, their significance levels and the R^2 from the regressions.

The estimated coefficients are highly significant for most variables and countries. Especially the effects of books at home and of parents' education are of a high magnitude. The sign of the effects is as expected for all variables. The only (marginally) significant counter-intuitive effect is a negative coefficient for a favorable attitude of parents in Iceland. The R-squared varies between 11 percent in Iceland and 29 percent in Hungary.

8. The results are based on the original scores. Selected results are also presented using the rescaled scores in the appendix. Further tables are provided by the author upon request.

Table 4.1
Estimates of student background effects for grade four (PIRLS)

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	16.83* (1.66)	11.50* (2.48)	10.42* (1.74)	20.33* (3.52)	8.23* (2.24)	20.20* (2.85)	13.47* (1.83)	18.23* (2.26)	7.52* (2.18)	19.31* (2.27)	22.08* (3.22)	27.36* (3.76)	12.87* (2.06)	22.63* (2.30)
Age	-9.53* (2.72)	-4.05 (2.52)	-15.42* (2.13)	31.15* (5.13)	-25.51* (2.17)	2.95 (4.52)	-17.14* (2.28)	22.43* (4.09)	12.52* (3.32)	-8.57* (2.39)	25.72* (5.14)	13.52‡ (7.19)	-5.62† (2.61)	2.20 (3.47)
Parents' origin	-1.68 (2.24)	-8.17† (3.50)	-12.35* (2.03)	-1.39 (3.08)	-6.01† (2.85)	-18.63* (4.03)	-12.42* (4.00)	-9.80† (3.90)	-17.00* (4.02)	3.30 (3.20)	-5.01 (4.69)	0.05 (4.03)	-15.53* (4.28)	-12.28* (2.56)
Language	-30.84* (3.10)	-24.27* (5.06)	-28.77* (2.67)	-42.21* (5.41)	-19.76* (3.45)	-10.96‡ (5.92)	-35.08* (4.73)	-29.94* (3.74)	-28.76* (6.32)	-16.93* (6.04)	-34.64* (5.72)	-47.20* (6.61)	-37.36* (8.04)	-27.62* (3.70)
Books at home														
26-100	2.31 (3.18)	15.24* (4.04)	13.55* (2.23)	-6.52 (6.23)	12.07* (2.71)	18.11* (3.86)	18.49* (4.28)	10.75† (5.26)	9.49* (2.99)	5.18 (5.67)	16.70† (7.72)	-3.08 (4.90)	7.96‡ (4.61)	11.93* (3.27)
101-200	15.60* (3.85)	3-3.75* (4.80)	28.56* (2.99)	18.16† (7.25)	20.96* (3.82)	27.26* (5.67)	25.00* (4.51)	29.21* (5.33)	25.49* (3.88)	12.69† (6.10)	27.84* (8.12)	19.27† (7.56)	15.83* (5.02)	16.03* (4.00)
>200	27.72* (3.88)	39.00* (4.54)	35.06* (2.86)	26.88* (7.11)	32.24* (3.94)	39.77* (4.69)	37.79* (4.52)	25.80* (5.19)	38.23* (3.56)	11.90‡ (6.38)	34.17* (8.10)	31.44* (6.32)	18.05* (5.44)	20.80* (4.17)
Parents' education														
Upper Second.	9.62‡ (5.36)	17.11‡ (8.99)	24.12* (2.73)	21.48* (5.24)	12.41* (2.93)	18.37* (4.70)	24.87* (3.51)	21.31* (3.66)	17.37* (2.86)	21.04* (5.12)	32.03* (8.02)	17.72‡ (9.85)	9.85 (7.29)	17.67* (4.43)
Post-Second.	21.02* (5.63)	35.68* (9.57)	18.29* (2.94)	34.69* (12.14)	26.32* (3.63)	34.21* (6.01)	29.00* (5.81)	33.02* (5.63)	12.09‡ (6.35)	30.99* (5.50)	- -	40.86* (9.70)	12.67‡ (6.93)	26.93* (4.83)
University	51.80* (5.73)	35.54* (9.41)	33.61* (3.25)	43.77* (4.90)	34.23* (3.83)	55.87* (5.58)	52.26* (4.42)	38.55* (4.63)	28.36* (3.31)	46.65* (6.83)	51.27* (8.10)	48.27* (10.46)	30.17* (7.18)	35.23* (4.81)
School location														
City	-1.29 (4.25)	12.75† (5.40)	-8.61† (4.23)	2.53 (6.66)	5.08 (4.41)	14.15† (6.27)	3.11 (3.44)	4.37 (4.09)	-2.29 (5.57)	8.25‡ (4.46)	16.26† (6.95)	15.86† (6.67)	4.79 (5.13)	-0.66 (5.67)
Rural	-14.32* (4.00)	0.89 (6.89)	-9.49 (9.48)	-3.81 (15.82)	7.70 (6.60)	9.45 (18.36)	-3.87 (4.73)	-6.47 (8.03)	-24.20† (12.08)	6.67 (9.34)	7.58 (5.41)	-10.27 (12.28)	- -	-0.91 (6.65)
Parents' attitude														
High	6.70* (2.28)	4.84‡ (2.91)	9.72* (2.04)	17.76* (4.80)	9.68* (2.53)	15.89* (3.86)	9.85* (2.31)	-5.06‡ (2.92)	14.95* (2.32)	10.77* (2.07)	9.56† (4.67)	26.54* (5.09)	5.39† (2.41)	6.23* (2.20)
Low	-4.54 (4.55)	-7.32 (9.01)	3.65 (2.76)	2.70 (8.66)	-5.47 (5.29)	7.22 (6.90)	-1.74 (5.78)	-12.72‡ (7.28)	-0.08 (4.74)	-1.48 (8.64)	-3.57 (6.18)	-11.33 (9.98)	5.27 (5.93)	-9.25† (4.67)
R-Squared	0.17	0.15	0.25	0.18	0.26	0.24	0.29	0.11	0.16	0.16	0.15	0.21	0.14	0.15

Coefficients from cluster-robust linear regressions (CRLR) by country, weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Significance levels: * 1 percent, † 5 percent, ‡ 10 percent. Dependent variable: PIRLS original reading literacy score.

The coefficients and their significance levels of the regressions using only the observations with no imputed values are presented in Table A4.6. Few coefficients change in terms of significance level or sign and no signs of significant coefficients change in comparison to

Table 4.1. The R-squared is higher in most countries for the regressions using only the original data, underlining the fact that imputed values of the explanatory variables contain less variation with which to explain the variation in test scores, as noted before. Table A4.7 additionally presents the estimates of student background effects from the school-fixed effects model including imputed data. For most variables and countries, the size of the effect is lower in absolute terms compared to the estimates in Table 4.1. The results show that student background has an effect both for the selection of students to schools and on student performance given the choice of school. The sum of these effects is of interest here and is presented in Table 4.1.

Effects at lower secondary education

Table 4.2 presents the estimates from the PISA study. Similarly to the former estimates, most coefficients are highly significant and are of the expected sign. Only the effect of parents' origin is significantly positive in four countries. The effect of a high parental attitude in Iceland is positive here.

Table 4.2
Estimates of student background effects for grade nine (PISA)

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	26.48* (1.42)	21.11* (3.52)	22.60* (4.31)	23.99* (4.45)	16.01* (2.40)	33.23* (3.59)	24.68* (3.51)	35.41* (3.29)	25.81* (4.29)	42.63* (3.89)	36.37* (3.44)	39.64* (4.38)	33.19* (2.76)	29.69* (2.55)
Age	-1.95 (2.72)	-27.44* (5.36)	-60.22* (7.19)	17.61* (4.80)	0.00 (3.36)	5.71 (5.33)	-35.24* (5.68)	7.61 (5.14)	11.83* (4.36)	-27.90* (7.17)	19.27* (5.40)	29.14* (5.21)	-13.35† (5.19)	15.70* (4.29)
Parents' origin	2.74 (2.36)	7.47† (3.63)	20.80† (12.56)	2.44 (5.03)	-6.98* (2.79)	5.56 (5.98)	-5.04 (6.22)	-8.88 (7.38)	16.21* (5.14)	-8.01 (5.49)	-9.24† (4.18)	6.53† (3.92)	3.11 (3.58)	0.79 (3.46)
Language	-34.15* (2.83)	-16.73 (10.17)	-66.20* (13.33)	-45.67* (8.07)	-22.85* (3.54)	-38.84* (9.92)		-29.12† (10.86)	-30.38* (5.19)	-25.93† (11.79)	-30.68* (5.17)	-82.18* (5.19)	-4.42 (11.96)	-43.29* (6.33)
Books at home														
51-100	20.12* (2.24)	24.25* (3.45)	36.76* (5.79)	27.60* (3.72)	20.67* (2.61)	22.00* (3.32)	24.31* (3.58)	15.76* (5.36)	17.01* (3.68)	7.49 (6.70)	22.67* (5.75)	35.97* (4.74)	16.83* (3.32)	24.75* (4.07)
101-250	30.27* (2.34)	40.01* (3.73)	61.87* (7.29)	52.24* (4.21)	33.18* (3.09)	33.32* (4.37)	46.30* (3.83)	33.89* (4.86)	28.56* (4.43)	32.24* (5.82)	51.08* (5.62)	57.04* (4.76)	44.04* (3.72)	38.77* (3.78)
>250	38.99* (2.41)	60.09* (4.27)	76.25* (8.09)	63.13* (4.71)	39.26* (3.50)	47.77* (5.91)	60.77* (4.66)	47.65* (4.39)	33.07* (5.18)	42.48* (6.09)	54.09* (5.50)	68.15* (5.25)	49.44* (4.24)	61.49* (3.56)
Parents' education														
Upper Second.		4.66 (10.50)	14.57† (7.27)	9.43 (5.19)	8.79† (3.45)	17.71* (6.37)	24.90* (6.01)	15.03* (4.97)	12.44* (4.03)	12.12 (13.33)	19.25* (6.60)	19.39* (7.06)	23.86† (10.44)	12.78† (6.28)
Post-Second.		28.06* (10.36)	39.38* (7.24)	50.22* (6.95)	14.17* (3.71)	29.75* (4.10)	53.12* (6.37)	22.88* (4.42)	20.99* (3.47)	19.71† (8.12)	9.69 (6.63)	40.11* (7.56)	35.71* (7.18)	18.38* (5.12)
University	19.07* (1.62)	61.18* (10.86)	40.32* (7.02)	29.11* (5.34)	8.13† (3.35)	38.13* (4.49)	78.73* (7.17)	25.57* (4.70)	26.78* (4.13)	29.48* (8.60)	13.61† (6.33)	29.85* (7.15)	43.07* (8.01)	11.86† (5.21)

Table 4.2 continued
Estimates of student background effects for grade nine (PISA)

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
School location														
City	-	-0.47 (6.63)	-10.57 (12.66)	0.83 (6.93)	5.00 (6.21)	14.66‡ (8.65)	19.94* (7.03)	-	3.25 (8.71)	15.19‡ (8.99)	4.13 (6.43)	19.12* (5.98)	18.60† (7.739)	3.44 (5.73)
Rural	-	-23.75* (7.32)	-16.65‡ (9.49)	-1.55 (6.96)	-0.62 (7.46)	-12.01 (13.36)	-26.42 (16.75)	-	10.56 (14.21)	-28.14* (8.57)	-13.53* (5.17)	-11.88‡ (7.08)	-19.64† (8.98)	-4.99 (3.85)
Parents' attitude														
High	22.49* (2.17)	9.51† (3.81)	16.34* (5.41)	31.24* (4.27)	8.84* (2.17)	6.01 (3.93)	7.31* (2.67)	14.26* (4.26)	1.68 (3.28)	11.01* (3.93)	25.75* (4.81)	15.13* (5.33)	12.27* (2.63)	13.10* (4.11)
Low	-27.02* (1.96)	-20.72* (2.60)	-9.14 (5.92)	-25.00* (3.61)	-13.79* (2.99)	-22.53* (4.78)	-13.29* (4.01)	-30.01* (4.42)	-23.50* (4.47)	-16.01* (5.01)	-38.92* (3.76)	-19.94* (4.34)	-15.55* (2.86)	-33.42* (2.99)
R-Squared	0.21	0.31	0.41	0.20	0.51	0.25	0.42	0.15	0.28	0.28	0.19	0.25	0.23	0.23

Coefficients from cluster-robust linear regressions (CRLR) by country, weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Dummies for grade levels are included in regressions. Significance levels: * 1 percent, † 5 percent, ‡ 10 percent.

Dependent variable: PISA original reading literacy score.

The R-squared is consistently higher in all countries compared to the regressions for primary education, which, to the extent that the measurements in the two studies are comparable, implies that student background has more of an effect on student test scores in lower secondary than in primary education relative to other factors like innate ability. The R-squared ranges from 15 percent in Iceland to 52 percent in France.

Table A4.8 displays the coefficients for the original data without any imputed values. All observations with missing values have been dropped. For some of the estimates the significance level or sign changes compared to Table 4.2 but no signs of significant coefficients change. The estimates from the school-fixed effect model are presented in Table A4.9. For most variables and countries, the size of the estimates is again smaller once we control for school choice. The decrease in the size of the estimates is mostly larger at grade nine than at grade four. Hence, selection into schools seems to be more important at lower secondary education than at primary education.

Comparison of results

In the following, the results from the previous section for the two studies are compared to each other. It is shown how the effects of student background variables on student performance – i.e. the educational opportunities – change between the end of primary and lower secondary education.

Table 4.3
Changes in the absolute size of effects between grade four and grade nine

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	+	+	+		+	+	+	+	+	+	+	+	+	+
Age	-	+	+	-	-		+	-		+		+	-	+
Parents' origin		-	+				-		-	+			-	-
Language			+			+						+	-	+
Books at home														
26-100	+	+	+	-	+							-	+	+
101-200	+		+	+	+		+			+	+	+	+	+
>200	+	+	+	+			+	+		+	+	+	+	+
Parents' education														
Upper Secon.														
Post-Second.			+		-		+						+	
University	-	+		-	-	-	+	-			-			-
School location														
City							+							
Rural		+							+	+	+			
Parents' attitude														
High	+			+				+	-		+		+	
Low	+		+	+		+		+	+		+		+	+

A + indicates an increase, a - a decrease in the absolute size of the effect from PIRLS to PISA. All changes in the effects are significant at the 10 percent-significance-level. Results are based on original scores.

A plus in Table 4.3 indicates an increase that is significant at the ten percent-significance-level in the absolute size of the coefficients β_1 and β_2 in equation (4.1) from PIRLS to PISA using the original scores. Looking at the characteristics of students, the sex of students has a significantly stronger impact on student reading test scores in all countries except for England in PISA than in PIRLS. Girls have expanded their advantage compared to boys at grade nine compared to grade four. Changes in the effect of age are not consistent across countries.

The following variables indicate the social and cultural background of students. The mostly negative signs for parents' origin imply that a foreign born parent has a less negative or even positive impact on student test scores at grade nine relative to grade four. However, speaking a different language at home has a more negative effect in four countries. The impact of the number of books at home increases with student's age. This finding is consistent across almost all countries. Parents' education has less impact on the test scores of 15-year-old students in seven countries, while the effect has increased in four countries. The absolute effect of community location increased in some countries. A rural location has a stronger negative effect in four countries. Changes in the effect of parents' attitude are fairly consistent across

countries and depict a stronger impact on student test scores. Table A4.10 presents the same evidence using the rescaled scores, i.e. the average score and standard deviation for the sample of 14 countries is identical for PIRLS and PISA. Overall, there are fewer significant changes but the sign of the changes is almost identical compared to Table 4.3. The school-fixed effects estimates differ less often significantly between PIRLS and PISA as well. For PIRLS, the difference between the total effect of student background and the effect conditional on school choice is much smaller than for PISA, where the school-fixed effects estimates are much lower in absolute size. Hence, the increase in the absolute size of the estimates is lower for the school-fixed effects estimates than for the total effect of student background.

The results show that the attitude and interest of students and parents become more influential (e.g. language spoken at home, parents' attitude), while family characteristics are losing their importance (e.g. parents' origin and education). Therefore, educational opportunities seem to increase for individuals with a positive attitude towards education. It is questionable in how far the number of books at home represents parents' interest in education or is an indicator of social background or both.

As the two studies are not directly comparable, it remains unclear what part of the changes in educational opportunities is due to differences between the studies and what changes really take place, though. The difference in the test design as well as the scaling of the test scores might cause a systematic greater dependence of test scores on student background in PISA compared to PIRLS. Even then, the result that the structure of educational opportunities changes should hold because the systematic difference should affect all student background variables. This is supported by the evidence based on the rescaled test scores. The comparison of the two studies in section 4.2.2 showed that the definition of the background variables might differ between the studies, especially for parents' education. The seemingly stricter definition of university education in PIRLS compared to PISA might at least partly explain the estimated negative signs in Table 4.3.

4.4 The role of institutions

This section aims at establishing the link between the institutional setting of the schooling systems and the former results on changes in educational opportunities in the countries. First, the schooling systems of the different countries are described according to certain criteria. Second, theoretical considerations on the link between schooling systems and educational opportunities are pointed out in relation to section 2.1 before the estimation strategy and the empirical evidence is presented.

4.4.1 *Schooling systems*

Although all countries except for Canada and New Zealand are European countries, their cultural background and population as well as their institutions may differ to some degree. Moreover, educational expenditure from the public and private side may differ. According to

the general cultural background, the former socialist countries CZE, HUN, LVA and RUS, the Scandinavian countries NOR and SWE and the Western European countries DEU, FRA and ITA are likely to feature similar schooling systems. In the following, the countries are described by the criteria schooling institutions, educational expenditure, and country and population parameters.

Table A4.18 in the appendix presents some features of the national schooling systems and of educational expenditure as well as country and population facets. Although there is no obvious link between the general cultural background and the institutional setting, the figures point at similarities within the groups of countries mentioned above. Streaming is measured only by the number of school types, i.e. by between school streaming and not within school streaming. Between school streaming is more likely to affect the impact of social background on performance to the extent that a social differentiation of students takes place. The degree of streaming is assumed to increase with the number of school types there are in the lower secondary schooling system. Compulsory education is organized in a single structure system in the Scandinavian countries and in ISL and the PISA index of school autonomy is comparatively high, which is consistent with the Nordic “local control” model of educational control (cf. Green et al., 1999). The PISA index indicates the freedom of choice of the heads of school in financial and personnel decisions.

In CZE, HUN and RUS lower secondary education is differentiated into two school types. Besides the single structure schools which are attended by most of the students, a low but growing number of students attends separated secondary schools (*gymnazium*) (cf. Anweiler et al., 1996). The PISA index of school autonomy is high in CZE and HUN as well as the private education expenditure as share of total education expenditure. However, the rate of students in private schools is rather low in these countries. This ratio is remarkably high in France (21 percent), whereas the relative private education expenditure in France is about average. The high German figure is due to the vocational training in the dual system. Due to data availability we only distinguish between publicly operated and funded schools and other schools. In general, the parameter values of the Western European countries DEU, FRA and ITA do not point at similar schooling systems, in terms of e.g. the number of school types or the total intended instruction time.

The dispersion of educational expenditure is relatively low. The expenditure on educational institutions as a percentage of GDP is higher than average in the Nordic countries, in France and in New Zealand. The country and population data indicate a higher female employment rate and also a tendency to a higher share of foreign population in countries with a relatively high rate of urban population.

4.4.2 Theoretical aspects

This part describes the expected links between educational opportunities and the schooling institutions. It summarizes the discussion of the literature and the results of the theoretical model presented in section 2.1. To discuss the link between background factors and institu-

tions, the variables are aggregated to measures of social background and parental attitude, similar to the distinction in the theoretical model.

The social origin of students is indicated by the number of books at home and parental education. Social differentiation might be both reduced and increased by the use of streaming in lower secondary education and the size of the private school sector. The direction of the link depends on whether school types for low achievers succeed in supporting these students and whether mobility between school types and public and private schools is sufficiently high. The degree of streaming in a schooling system is indicated by the number of major school types. The more school types there are the greater is the differentiation of the system. This implies that the restriction has been imposed that the effect of three school types is twice the effects of two school types. An alternative measure of streaming is a binary variable on whether there is more than one school type or not. The theoretical model predicts higher inequality through the use of streaming and a large private school sector. A higher instruction time instead should be linked to a lower inequality in educational opportunities.

The impact of parents' attitude towards their child's educational success may be linked to several factors. A high instruction time might limit the influence of parents, while a higher autonomy of secondary schools may increase the possible effect on student test scores because it is easier for parents to influence local decision makers like teachers than some regional bureaucrats. This is also the result of the theoretical model.

4.4.3 Estimation strategy

Two different approaches to assess the impact of schooling institutions on the degree of equality of educational opportunities will be applied in the analysis. The first approach is straightforward and directly relates the institutional variables to the impact of student background factors on test scores for each cross section of PIRLS and PISA separately.

To this purpose we repeat equation (4.1) with subscript $t = 4$ or 9 , referring to PIRLS or PISA, respectively, and a subscript y referring to country y and suppressing the subscript s referring to schools:

$$(4.2) \quad T_{tyi} = \beta_{ty0} + B_{tyi}\beta_{ty1} + D_{tyi}\beta_{12} + \varepsilon_{tyi}$$

and we specify:

$$(4.3) \quad \beta_{ty1} = \beta_{t3} + A_t I_{ty} + c_{ty}$$

where I_{ty} is a vector of institutional variables in country y relevant at primary ($t = 4$) and secondary school ($t = 9$) level, the parameters in the matrix A_t represent the effects of institutional variables on the parameters of (some of) the background variables and c_{ty} denotes unobserved country (cultural) effects. To be specific, for the primary school effects:

I_{4y1} : yearly instruction time in primary school in minutes. Is assumed to affect the parameters of all four selected background variables. No data available for Hungary.

I_{4y2} : share of students in private education in primary education. Is assumed to affect the parameter of the highest category of number of books and the highest category of parents' education.

For the secondary school effects:

I_{9y1} : yearly instruction time in secondary school in minutes. Is assumed to affect the parameters of all four selected background variables. No data available for Norway.

I_{9y2} : share of students in private education in secondary education. Is assumed to affect the parameter of the highest category of number of books and the highest category of parents' education.

I_{9y3} : number of school types in secondary education to indicate the degree of streaming (between 1 and 3, see section 4.4.2; alternative measure: 1 if more than one type of secondary school, 0 otherwise). Is assumed to affect the parameter of the highest category of number of books and the highest category of parents' education.

I_{9y4} : PISA index of school autonomy. Is assumed to affect the parameters of parents' attitude. No data available for France and Latvia.

Substituting (4.3) in (4.2) yields:

$$(4.4) \quad T_{tyi} = \beta_{ty0} + B_{tyi}\beta_{t3} + B_{tyi}A_t I_{ty} + D_{tyi}\beta_{t2} + \varepsilon_{tyi} + B_{tyi}c_{ty}.$$

Using the sample of 14 countries, estimates can be obtained for the parameters of interest, contained in the matrices A_4 and A_9 . As a result of the introduction of country dummies (in view of the country-specific constant term) the effect of the left-out cultural heterogeneity $B_{tyi}c_{ty}$, may be expected to be mitigated, but will not be annihilated. If the cultural effects are correlated with schooling institutions and determine educational opportunities this could potentially render some or all of the coefficient estimators inconsistent.

The second approach makes use of the within country variation over student age by using both the PIRLS and PISA data. Thereby, a strategy can be employed that eliminates country effects under the assumption that they are constant for students in grades 4 and 9. This approach is more closely related to the recent literature on streaming and educational opportunities that uses the difference-in-differences methodology (Hanushek and Woessmann, 2006; Waldinger, 2006). These authors argue that cultural effects are likely to bias the estimator in a cross-sectional estimation approach and regard difference-in-differences as the preferred estimation approach. However, both approaches need assumptions to identify the effect of schooling institutions on educational opportunities. Hence, the assumptions are explicitly stated and the results from both approaches are presented here. Nevertheless the results of the first approach are presented in the appendix.

Determining the role of institutions in international comparisons of schooling systems depends crucially on the assumption that one can control for country-specific differences, such as cultural and social factors. It is however doubtful that traditional behavior and atti-

tudes towards education can be grasped by standard variables. Instead, the strategy chosen is to eliminate country-specific factors. First, not the level of educational opportunities but the changes between grade four and grade nine in educational opportunities are estimated for each country. In a second step, these changes are related to changes in institutions of the schooling systems. The intuition behind the estimation strategy is the following. When we consider the relationship between the share of students that attend private schools and the effect of parents' education on student test scores, we would expect a higher impact of parents' education in countries in which we observe a large private school sector, assuming that private schools provide better schooling for students from a higher social background (cf. section 2.1). Relating the effect of parents' education to the size of the private school sector in lower secondary education would only reveal this relationship if there were no other differences between countries that could explain the size of the effect of parents' education on student test scores. Therefore, the estimation strategy relates the changes in the effect of parents' education between grade four and grade nine to the changes in the size of the private school sector.⁹ In countries in which the size of the private school sector increases strongly between the two grades, we expect that the effect of parents' education on student test scores increases as well (see section 2.1). However, in countries in which the importance of private schools is the same in grade four and grade nine, there should be no change in the effect of parents' education, irrespective of the absolute size of the private school sector. The details of the second estimation strategy are presented more formally below.

The first problem we face is that the two studies on which the analysis builds are not identical and the skills tested might differ slightly. Taking the PIRLS study as the reference study, the PISA test score is only an imperfect measure of what students would have achieved in a PIRLS-like test at grade nine. We assume that the measures of student background are identical in PIRLS and PISA, which use very similar questionnaires. The test score in PISA T_{9yi} , however, is only an imperfect measure of the 'true' test score. There is hence a measurement error e_{9yi} . The production function then can be written as:

$$(4.5) \quad T_{9yi} = \beta_{9y0} + B_{9yi}\beta_{9y1} + D_{9yi}\beta_{92} + e_{9yi}.$$

The subscript 9 implies that students are tested in grade nine, y indicates the country and i the individual. The index for schools and the school effects are omitted for simplicity. Under the assumption that the measurement error e_{9yi} is not related to student background B_{9yi} and the vector D_{9yi} , the estimator for β_{19y} is unbiased.¹⁰

The estimated coefficient β_{9y1} is then explained by the institutional setting in country y as in equation (4.3). This or a similar equation is often estimated in other studies (e.g. Schuetz et al., 2005) and by the first approach, where c_y is either omitted or replaced by proxy variables

9. 'Grade nine' refers to the PISA data and actually includes all 15-year-old students in grades eight, nine and ten. The estimations were also run only for students in grade nine to check for a possible selection bias due to the PISA sampling design (see sections 2.2 and 4.4.5).

10. For a more formal presentation of the assumptions concerning the difference the PIRLS and the PISA study, see Ammermueller (2005).

such as GDP per capita. However, it is likely that schooling institutions are strongly linked to cultural background which can hardly be grasped by proxy variables, i.e. $\text{Cov}(I_{ty}, c_{ty}) \neq 0$. In this case the estimator of A_i is biased and we cannot observe the relationship between institutions and educational opportunities. In order to overcome this problem of omitted variables, we explain the differences in coefficients instead of the level of coefficients by changes in institutions. This approach also controls for a difference in the correlation between student ability and student background across countries, as long as this correlation does not change over student age. Taking differences of equation (4.3) between grade 9 and grade 4 at the country level leads to the following equation:

$$(4.6) \quad \beta_{9y1} - \beta_{4y1} = \beta_{93} - \beta_{43} + A_9 I_{9y} - A_4 I_{4y} + c_{9y} - c_{4y}.$$

We now assume that the effects of the institutional factors I_{ty} and the unobservable country-specific effects c_{ty} on educational opportunities do not change between grade four and grade nine, i.e. $A_4 = A_9$ for the non-zero elements in the matrices and $c_{4y} = c_{9y}$. The latter assumption is the key identifying assumption of the second approach and is the standard difference-in-differences (DiD) assumption (also see Meyer (1995) for a discussion of the assumptions in a DiD setting). Waldinger (2006) regards this assumption as less restrictive than the assumption of the first approach, $\text{Cov}(I_{ty}, c_{ty}) \neq 0$. The first assumption applies only to the institutional variables for which there is variation at both grades (instruction time and the share of students in private schools) and is necessary to have sufficient degrees of freedom at the country level to identify all parameters. Under these assumptions, we obtain:

$$(4.7) \quad \beta_{9y1} - \beta_{4y1} = \beta_5 + A(I_{9y} - I_{4y}),$$

where the matrix $A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & 0 & 0 \\ \alpha_{31} & \alpha_{32} & 0 & 0 \\ 0 & 0 & \alpha_{43} & \alpha_{44} \end{bmatrix}$ equals the matrix A_9 described above.

Each element of matrix A represents an interaction between the four student background variables and four institutional variables described above. The coefficient α_{12} states the interaction between parents' university degree and instruction time for example. The transposed matrix of interaction terms is depicted in Table 4.4, where the blank spaces indicate the zero coefficients in matrix A .

Moreover, we have to assume that no anticipation effects take place, i.e. that institutional features in lower secondary education take effect already in primary education. For one institutional feature, the degree of school autonomy, information is only available for grade nine. Here we have to assume that school autonomy is identical across all countries at grade four. In the estimation of equation (4.7) school autonomy at grade four is set to zero. Hence, we can only exploit the variation in school autonomy at grade nine.

In most countries considered in this analysis, the specific features of the schooling system like streaming or a high share of private schools are introduced only after primary education. If this holds for the degree of school autonomy as well, the assumption that school autonomy is identical across countries at grade four should hold. Under the assumption that the impact of country-specific effects does not change between the fourth and ninth grade or that the change is not related to the institutional factors, country-specific factors are largely eliminated by this estimation strategy. The results should be interpreted as correlations, however, because the interaction effects may be determined by other observable and unobservable institutional variables as well, so that the estimator of A may be biased due to omitted variables. This applies even more to the estimations by Hanushek and Woessmann (2006) and Waldinger (2006). Unfortunately, this data demanding estimation strategy leads to a low number of observations in the second step, which leaves little room to explain the interaction effects by several institutional features.

The second approach can be either estimated in a two-step model or a pooled model. In the two-step model the β_{4y1} and β_{9y1} are estimated first using equation (4.2). The second step then estimates equation (4.7). Details of the estimation procedure of the two-step model are given below, which is used in particular to show the robustness of the results. Writing equation (4.2) as a pooled model including both PIRLS and PISA data with P being a dummy variable that is one for PISA and zero for PIRLS yields:

$$(4.8) \quad T_{yi} = \delta_{y0} + P_{yi}\delta_{y1} + D_{yi}(\delta_{y2} + P_{yi}\delta_3) + B_{yi}(\delta_{y4} + P_{yi}\delta_5) + \varepsilon_{yi}$$

where $\delta_{y4} = \beta_{4y1}$ and $\delta_{y5} = \beta_{9y1} - \beta_{4y1}$ from equation (4.2). Note that δ_{y2} is indexed by country in equation (4.8). Substituting equation (4.7) into equation (4.8) leads to the pooled model:

$$(4.9) \quad T_{yi} = \delta_{y0} + P_{yi}\delta_{y1} + D_{yi}(\delta_{y2} + P_{yi}\delta_3) + B_{yi}(\delta_{y4} + P_{yi}\delta_5) + B_{yi}P_{yi}A(I_{9y} - I_{4y}) + \varepsilon_{yi}.$$

Here, all students i from all countries y from both studies are pooled in one regression. The subscripts for schools and school effects are omitted. P_{yi} is a dummy that is one for observations from the PISA data and zero otherwise. The country-specific intercepts for PIRLS are denoted by δ_{y0} , while $\delta_{y1}P_{yi}$ allows for country-specific intercepts for PISA. The coefficients δ_{y2} and δ_{y4} measure the country-specific effects of student background, while δ_3 and δ_5 indicate the change in the impact of student background from grade four to grade nine. Notice that δ_3 and δ_5 are not indexed by country. The cross-country variation that is spared is used to identify selected interaction effects between the student background variables of interest and institutions. The matrix A has been introduced above. The non-zero coefficients in matrix A express the difference in the coefficient of the student background variable resulting from a change in the respective institutional variable that is part of vector I between PIRLS and PISA by one unit. As in the two-step estimation approach, it is assumed that changes in educational opportunities vary with certain features of the schooling institutions. In the pooled model, all interaction effects can be estimated within one model. Moreover, the pooled model includes country-specific effects of student background for one

study (PIRLS) and is hence less restrictive than the two-step model or the pooled model used by Waldinger (2006).

The identification of the institutional effects in the second approach is driven by the difference in the institutional setting to which students are exposed at the end of primary and lower secondary education. Many countries move from a comprehensive primary school system to streaming in secondary education or extend the amount of instruction time or the size of the private school sector. If students in primary school and their parents are able to anticipate the changes waiting for them in secondary education, this would lead to a bias of the estimator of the components of A towards zero. However, it is unlikely that anticipation strongly limits the impact of the current institutional setting, because students and parents are not offered any alternatives.

Equations (4.4) and (4.9) are estimated by an efficient feasible generalized least squares (FGLS) estimator with a heteroskedastic error structure. Equation (4.9) is estimated as well using weighted cluster-robust linear regressions (CRLR) to allow for a possible dependence of error terms of observations within clusters (White, 1980, p. 821; Deaton, 1997, pp. 73-78; see section 2.2.2). The schools and not the countries are chosen as clusters because country-specific intercepts for both PIRLS and PISA and country-specific effects of student background are included in this pooled model. This should be sufficient to control for heterogeneity in the effects of observed characteristics across countries and unobserved country-effects. In order to check the robustness of the results with regard to a possible aggregation bias, estimates using countries instead of schools as clusters are discussed in section 4.4.5 as well in addition to the results of the two-step model.

4.4.4 Empirical evidence

In order to test the hypotheses stated above, the results from both estimation approaches are discussed. First, the results of the cross-sectional estimation strategy based on equation (4.4) for both PIRLS and PISA are presented. Second, the second approach is applied using both the pooled model presented in equation (4.9) and the second step of the two-step model shown in equation (4.7). The features that are used to describe the institutional setting include the amount of instruction time, the share of students that attend private schools, the number of school types and an index of autonomy of lower secondary schools. They are presented in Table A4.18 in the appendix. Recall that there is no variation in the number of school types and no information on school autonomy in primary education.

Tables A4.11 and A4.13 present the results of a FGLS estimation from the cross-sectional approach for PIRLS and PISA, respectively. For PIRLS, all interaction effects are significantly positive. Positive coefficients have been expected from the theoretical hypotheses for the interactions involving the share of students in private schools and the interaction between instruction time and a low parental attitude. The results for the further interaction effects for the amount of instruction time oppose the expected effect. For PISA, also the effects of the number of school types and the degree of school autonomy can be examined. Six out of

nine significant coefficients support the hypotheses. The share of students in private schools decreases instead of increases the impact of social background and instruction time is only negatively linked to the effect of books at home but not parental education. The coefficients in Table A4.11 and A4.13 are estimated consistently only under the assumption that there are no country-specific effects that are correlated both with the impact of student background on test scores and schooling institutions (cf. section 4.4.3). Under the assumption of the absence of country-specific effects we can test the assumption that schooling institutions have the same impact on educational opportunities at grade four and grade nine ($A_4 = A_9$ in equation 4.6). Table A4.15 presents the test statistics and significance levels of a Wald test on the equality of coefficients between PIRLS and PISA. The test results show that the hypothesis of equal coefficients can be rejected in four out of six cases. Hence, only for two interaction effects the hypothesis seems to be justified. However, both the cross-sectional results and the test results are only consistent in the case of absence of any country-specific effects. As was argued before, this is a strong assumption which is unlikely to hold and has been criticized in the literature as well (Hanushek and Woessmann, 2006; Waldinger, 2006). When the cross-sectional estimations are performed using a school-fixed effects model, the results are fairly similar (cf. Tables A4.12, A4.14, A4.16)

The following two tables now present the estimates from the second approach of the coefficients from the matrix A from equation (4.9), in which the changes in educational opportunities are interacted with the changes in institutional variables and are estimated in a pooled regression including all 14 countries.¹¹ Only the interaction effects which test the stated hypotheses are included in the regression. Table 4.4 presents the estimates from the efficient FGLS estimator and Table 4.5 from the more robust cluster robust linear regressions (CRLR). The coefficient of 1.29 indicates that the impact of having more than 200 books at home increases by 1.29 test score points from grade four to grade nine when the share of students in private schools increases by one percentage point between primary and lower secondary education. A higher instruction time is associated with a lower impact of social background as indicated by the number of books at home. The results indicate a positive relationship between the impact of parents' education and the number of school types. A higher degree of school autonomy is positively linked to the impact of a high parental attitude towards their child's education. The estimated coefficients hardly differ between the FGLS estimates in Table 4.4 and the CRLR estimates in Table 4.5 while the standard errors of the FGLS estimates are smaller. The only major difference in the results concerns the significance level of the estimates involving the share of students in private schools. While the FGLS estimate for the interaction with parents' education is significant at the five-percent level the corresponding CRLR estimate is highly insignificant. When the equations are re-estimated using only one indicator for social background, either the number of books at home or parents' education, the results are qualitatively the same.

11. Since the information on institutions is not always available for all countries (see Table A4.18), separate regressions were performed for those institutional variables with missing values, always including all possible interaction effects. In the regressions for the institutional variables number of school types and private schools, 151,542 students in 14 countries, for instruction time 134,635 students in 12 countries and for school autonomy 136,541 students in 12 countries are included.

Table 4.4
Results from pooled model FGLS

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.002* (.0004)	1.29* (.46)	-.55 (1.24)	
One parent univ. degree	.0003 .0004	-.97† (.45)	9.86* (1.28)	
Parents' attitude				
High	.0006 .0004			3.56* (.67)
Low	.0002 (.0006)			.86 (1.10)

Interaction effects between changes in institutions and changes in student background effects are presented. Estimated using equation (4.9) by FGLS. Canada and England are included in the regressions. Dependent variable: PIRLS/PISA original reading literacy score.

FGLS estimations with heteroskedastic panels.

Significance level: * 1 percent. † 5 percent. ‡ 10 percent.

Although the sign of the effects can be reasonably interpreted, it is difficult to compare and interpret the size of the effects due to possible differences between PIRLS and PISA test scores. Nonetheless, a “naïve” comparison is made between two countries to compare the size of the effects with the differences in student background effects across countries. Following the analysis in Chapter 3, Germany is compared to Sweden as an example of a Scandinavian country. Finland did not participate in PIRLS and for Norway no information on instruction time is available for PISA. The two indicators of social background, the number of books at home and parents' education, and the significant interaction effects with the three institutional features are considered. In PISA, both background factors have a significantly higher impact in Germany than in Sweden. The difference between the effects in PIRLS and PISA amounts to 41.19 points for the highest category of books at home and 6.71 for the highest category of parents' education in Germany and to 40.69 and -23.37 points in Sweden, respectively (see Tables 4.1 and 4.2). The increase in annual instruction time between primary and lower secondary education is higher in Germany than in Sweden and can explain -3.23 points of the 0.5 point difference between Germany and Sweden from the differences in the effect of books at home between PIRLS and PISA according to the results in Table 4.4. The difference in the increase in the share of students in private schools adds 2.84 points. For the effect of parents' education, the institutional factor private schools contributes -2.13 points while school types with 19.72 points can explain a large part of the difference of 30.08 points between the two countries. This example shows that differences in schooling institutions between primary and lower secondary education cannot explain the total difference between the differences in student background effects between countries but a large share of it, in particular the degree of streaming.

Table 4.5
Results from pooled model CRLR

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.002† (.0007)	1.46‡ (.77)	2.22 (2.29)	
One parent univ. degree	-.0008 (.0006)	-.60 (.64)	8.63* (2.09)	
Parents' attitude				
High	.0006 (.0007)			3.53* (1.12)
Low	.00002 (.001)			-.04 (1.78)

Interaction effects between changes in institutions and changes in student background effects are presented. Estimated using equation (4.9) by CRLR. Canada and England are included in the regressions. Dependent variable: PIRLS/PISA original reading literacy score. Cluster-robust standard errors in parentheses. Schools are clusters. Significance level: * 1 percent. † 5 percent. ‡ 10 percent.

Table 4.6
Results from pooled SFE model (FGLS)

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.005* (.0004)	-.09 (.45)	-18.14* (1.22)	
One parent univ. degree	-.001† (.0004)	.12 (.44)	1.45 (1.22)	

Interaction effects between changes in institutions and changes in student background effects are presented. Estimated using school-fixed effects model of equation (4.9). Canada and England are included in the regressions. Dependent variable: PIRLS/PISA original reading literacy score. FGLS estimations with heteroskedastic panels. Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.17 in the appendix presents the coefficients and standard errors from the two-step model. Equation (4.7) is estimated by seemingly unrelated regressions, taking account of the small-sample characteristics. Seemingly unrelated regressions are used to increase the efficiency in the estimation of the four equations with different explanatory variables, which are likely to be dependent on each other. As the outcome variable of the second step is estimated in the first step, the second step regressions are weighted by the inverse of the standard error (cf. Card and Krueger, 1994). The regressions include between 10 and 14 observations. The low amount of observations is clearly a drawback of the estimation strategy, which is very data demanding. This applies as well to the pooled model, which can only use the variation between countries to identify institutional effects. The results do not differ from the pooled model except for less significant effects.

The results of the second estimation approach show that schooling institutions affect the degree of equality of educational opportunities. From a policy point of view, it is of interest to know which factors or processes drive these results. A possible factor is the selection of students into different schools. This is the main argument for the negative impact of streaming and private schools on the degree of equality of educational opportunities in the literature and the theoretical model. The impact of selection of students into schools can be assessed by estimating equation (4.9) by FGLS as a school-fixed effects model (SFE), i.e. including separate intercepts for each school. By controlling for differences between schools, the student background effects β_j in equation (4.1) do not include the effect of background that take place through the selection of students into different schools (see discussion of student background effects in section 3.3.1). We would hence expect that the institutional variables that describe the differentiation of the school systems, i.e. the share of students in private schools and the number of school types, have a lower or no impact on the effect of student background on test scores. The variables on school location and parents' attitude, which vary only at the school or class level, are not included in the SFE model. Table 4.6 presents the results. Indeed, three of the four coefficients of the institutional variables private schools and school types decrease compared to the previous results and the only significant interaction effect is a negative coefficient for the pair books at home / school types. The interaction terms for instruction time are instead both significantly negative as expected from the theoretical model. The impact of student background on test scores conditional on the school is hence still related to the amount of instruction time but not to the differentiation of the schooling system as indicated by the share of private schools and the number of school types, except for the counterintuitive negative impact of school types. These results confirm the arguments of the theoretical model, i.e. that selection into schools and possible school and peer effects lead to a decrease of the equality of educational opportunity.

Table 4.7 now compares the theoretical hypotheses on the link between educational opportunities and the institutional setting and the empirical evidence from the second estimation approach as well as that from the SFE approach. The expected effects are either supported by the empirical evidence or the empirical evidence is ambiguous, i.e. the effects are insignificant, with only one exception. The pooled model estimated by CRLR and the two-step model always lead to the same sign of the effect whenever the estimates are statistically significant. The pooled SFE model has the expected effects for instruction time but not for school types and private schools because it controls for all differences between schools. The results of the cross-sectional estimation approach, which have been discussed above, are instead counterintuitive in several cases and contradicting between the two cross sections.

Although precise inference is in many cases hampered by degrees of freedom, it is clear parents' education as an indicator of social background is positively linked to the number of school types. A greater choice among schools seems to benefit socially advantaged students, who might have easier access to higher school types. This also holds for the size of the private school sector, which is positively linked to the effect of books at home (but negatively to the impact of parents' education in the FGLS estimation). The influence of parents seems to be limited by the time students spend in school, which follows from the theoretical discussion.

Greater school autonomy is associated with a stronger absolute effect of positive parental attitude. Hence, school autonomy increases the possibility of parents to positively influence their child's school performance as measured by test scores.

Table 4.7
Comparing theoretical hypotheses and empirical evidence on institutional effects

Student background variable Institutional variable	Theoretical Hypotheses	Empirical evidence			
		Pooled FGLS	Pooled CRLR	Two-step	Pooled SFE
Books at home					
Total intended instruction time	Negative	Negative	Negative	Negative	Negative
Share of students in private schools	Positive	Positive	Positive	Positive	--
Number of school types	Positive	--	--	--	Negative
Parents' education					
Total intended instruction time	Negative	--	--	--	Negative
Share of students in private schools	Positive	Negative	--	--	--
Number of school types	Positive	Positive	Positive	Positive	--
Parents' attitude high					
Total intended instruction time	Negative	--	--	--	
Autonomy of schools	Positive	Positive	Positive	Positive	
Parents' attitude low					
Total intended instruction time	Positive	--	--	--	
Autonomy of schools	Negative	--	--	--	

Effects are significant at the 10 percent-significance-level. --: effects are insignificant. Theoretical hypotheses are based on sections 2.1.2 and 4.5.2, empirical evidence on Tables 4.4, 4.5, 4.6 and A4.17.

The results are in line with the evidence presented by Hanushek and Woessmann (2006). They show that streaming leads to an increase in test score inequality between primary and secondary education. Moreover, they provide evidence that this result is largely insensitive to the pairs of student performance studies that are used for the analysis. According to their results, streaming leads to lower or has no effect on mean student performance. The same effect of streaming is shown in the study by Brunello and Checci (2006) for the impact of family background on educational attainment and labor market outcomes. Waldinger (2006) finds a positive but insignificant impact of streaming on student background effects, using the timing of streaming as a measure. Schuetz et al. (2005) find no evidence of a significant relationship between instruction time and educational opportunities. However, they use a cross-sectional identification approach that controls for country-specific effects only to a lesser degree and take the length of the average school day instead of the amount of annual instruction time.

Table 4.8
Predicted effect of student background at grade nine

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Books at home														
>200	17.55*	39.98*	52.97*	37.78*	29.61*	32.70*	-	25.16*	23.11*	14.87†	-	46.59*	21.16*	59.00*
	(4.75)	(5.58)	(9.14)	(5.45)	(5.30)	(6.61)	-	(5.78)	(5.94)	(6.38)	-	(6.57)	(5.43)	(3.24)
Parents' education														
University	5.36	15.98	6.10	9.61	-13.28†	13.93†	-	-3.96	1.70	-0.49	-	9.10	-0.63	15.65*
	(5.35)	(9.45)	(7.20)	(5.89)	(5.55)	(6.02)	-	(6.25)	(5.59)	(7.55)	-	(8.28)	(7.18)	(4.17)
Parents' attitude														
High	14.71*	14.78*	8.05†	23.06*	-	7.94†	-	6.05	4.73	-	-	24.51*	4.74	17.52*
	(3.10)	(4.69)	(3.75)	(4.20)	-	(3.89)	-	(3.98)	(3.44)	-	-	(5.03)	(3.21)	(2.69)
Low	-13.82*	-8.80‡	1.00	-10.67‡	-	-8.07	-	-18.22*	-10.03‡	-	-	-9.73	-3.47	-32.37*
	(4.68)	(5.18)	(7.03)	(5.55)	-	(5.91)	-	(6.09)	(5.56)	-	-	(6.36)	(4.89)	(2.69)

Predicted student background effects for PISA and standard errors based on estimation of equation (4.9) by CRLR.

The estimates from the second approach can now be used to predict the impact of student background on test scores accounting for schooling institutions for PISA. The impact of the four student background variables presented in Table 4.8 is based on the estimations of equation (4.9) by CRLR (see Table 4.5). The presented student background effects are the sum of the country specific student background effects for PIRLS, the average change in the effect between PIRLS and PISA and the products between the estimates of institutional effects and the changes in the institutional variables between primary and lower secondary education. The presented values in Table 4.8 and the estimates of student background effects for PISA in Table 4.2 are highly correlated. The correlation coefficient is 0.83 for the number of books at home and 0.44 for parents' education. For a high parental attitude it amounts to 0.60 and for low parental attitude to 0.91. The correlation is much higher than between the 'total' student background effects for PIRLS and PISA, which ranges between 0.05 and 0.38 for the variables. The figures show that the variation in institutional variables and the estimation results contribute to explaining differences in educational opportunities across countries for PISA.

4.4.5 Robustness tests

Several alternative specifications have been estimated, using mainly the pooled model estimated by CRLR, in order to ascertain the robustness of the presented results.¹² First, all regressions have been repeated using only the original data and not the imputed values. The number of observations in the pooled model then drops from 151,542 to 86,792. The interaction effect between the number of books at home and instruction time turns insignificant.

12. Tables presenting the results for the robustness checks are available from the author upon request.

Second, Canada and England are excluded from the regressions of the pooled model because the sampled regions differ between PIRLS and PISA. The results are almost identical to those shown in Table 4.5, except that the interaction between school autonomy and a high parental attitude turns insignificant.

Third, the pooled model has been estimated using only students in grade four and grade nine or ten. Before, students from the PISA study could be in grade eight, nine or ten. The equations have been estimated using grade level dummies. However, depending on the school enrolment criteria, the grade level dummies might be endogenous in the equations and lead to biased estimators of educational opportunities. Therefore, only students from the most frequented grade level have been kept and the analysis was performed with this restricted sample. The results are again very similar and no changes in significance levels or signs occur.

Fourth, the discrete variable on the number of school types has been replaced by a dummy which indicates whether countries use streaming in lower secondary schooling or not. The school type variable was defined in such a way that it represents the degree of streaming that is applied in the countries. Of course, this is difficult to assess and may be subject to measurement error. The dummy variable allows for less variation between countries, but can be defined more clearly. The results strongly reinforce the presented evidence. The coefficients for both interactions with school types increase and both coefficients are significant at the one percent-level. However, the amount of instruction time is no longer significantly related to the amount of books at home.

Fifth, the estimations have been repeated using the rescaled scores for PIRLS and PISA instead of the original scores (see section 4.2.2). The size of the estimated effects changes slightly but the qualitative interpretation does not change. Only the link between books and private schools is no longer significant at the 10 percent-significance-level and the coefficient for books at home and instruction time moves to the ten percent-significance-level.

Finally, the aggregation bias of the estimator of the institutional variables that vary only at the country level is approached differently. Equation (4.9) includes country-specific intercepts for both PIRLS and PISA and country-specific covariates for PIRLS to control for heterogeneity across countries and country-level shocks. Equation (4.9) is then estimated using schools as clusters. One could argue that equation (4.9) does not control sufficiently for heterogeneity across countries and that standard errors for the institutional variables are underestimated due to aggregation bias (Moulton, 1990). Therefore, equation (4.9) is also estimated without country-specific covariates, i.e. δ_2 is not indexed by y , but using countries as clusters instead of schools. The pair books at home / school types turns significant while the pair parents' education / school types turns insignificant. The negative interaction term

between low parental attitude and school autonomy turns significant as expected from the theoretical considerations.¹³

Overall, the main results are robust in spite of the few observations on institutions (i.e. countries) and the rather imprecise definition of the variables characterizing them, like the degree of streaming. The links between books at home and instruction time as well as private schools are not supported by all robustness tests, as is the link between school autonomy and parental attitude. However, all robustness tests support the majority of significant effects and provide further evidence to underline the positive link between the impact of social background and the differentiation of the schooling system.

4.5 Conclusion

Intergenerational mobility of human capital is largely determined by educational opportunities of students. Therefore, creating equal opportunities should be a main aim for policy-makers and could prevent costly redistribution later on. Moreover, large inequalities in educational opportunities can lead to a sub-optimal investment in human capital. The schooling systems and their institutional frameworks are a key factor for promoting equality of opportunity in education. This chapter tries to explain cross-country differences in educational opportunities by analyzing the link between the institutional setting of schooling systems and educational opportunities. The estimation strategy, which exploits the information on changes in educational opportunities between two stages in a student's schooling career and differences in the institutional setting across countries and between primary and lower secondary education, controls largely for country-specific factors that invalidate other cross-country comparisons of institutional effects on student test scores. Overall, the estimation approach and specification of the model not only leads to further insights into the impact of other schooling institutions that have been neglected so far but should as well lead to a lower bias of the estimates compared to previous studies.

The empirical analysis builds on the international PIRLS and PISA studies on reading literacy. The estimation of the effects of student background on student performance shows that the absolute effect of gender, amount of books at home, school location and parents' attitude on student test scores increases between the end of primary and lower secondary education, while the effect of parents' origin and education seems to decrease. This implies that the attitude and interest of students and parents are becoming more influential, while family characteristics are losing their importance. Therefore, educational opportunities seem to increase for individuals with a positive attitude towards education.

13. Since the number of clusters is small, the t-statistics might not be normally distributed (Donald and Lang, 2004). For additional assumptions that have to be made in this case, see Wooldridge (2003). This critique does not apply to the former regressions using schools as clusters or the alternative two-step model. Alternative two-step results that interact each student background effect with each schooling institution in a separate second step regression are presented in Ammermueller (2005) and support the results.

It can be shown that institutions play a role in determining the educational opportunities of students. The evidence from the second approach on the link between institutions and changes in educational opportunities is in line with theoretical hypotheses, whenever it is significant. The social origin of students, measured by the number of books at home and parental education, increases its effect on student test scores in countries with a differentiated secondary schooling system with several school types and an increase in the size of the private school sector between primary and secondary education (not for all estimations). This supports the hypothesis that streaming and private schools benefit the performance of students as measured by test scores from a higher social background. The time students spend in schools seems to limit the effect of social origin upon test scores while higher school autonomy is associated with higher parental influence. Note that the direct impact of institutions on average student performance cannot be investigated by the approach chosen here. Recent empirical evidence suggests that the effect of streaming is not positive but rather negative (Hanushek and Woessmann, 2006; Maurin and McNally, 2006).

The results imply that schooling institutions can affect the equality of educational opportunities of students and have hence an impact on the intergenerational mobility of human capital. According to the findings, educational opportunities could be increased by a low differentiation of lower secondary education, a low share of private schools and a relatively high instruction time for example. School autonomy can have both a negative and a positive effect on student test scores, which depends on the attitude of parents. However, these suggestions are based on findings of conditional correlations and not of causal effects and should therefore be taken cautiously. Moreover, the effects of schooling institutions depend always on the implementation of policies within individual countries, which can hardly be measured. The comparability of different student performance studies as well as the selection of countries and their limited number cast some doubt on the generalization of the results. The link to the findings of the previous chapters is discussed in the following chapter.

5

Synthesis and conclusion

Human capital acquired by educational processes is an important determinant of both major individual and social outcomes. The benefits of education are manifold and range from higher earnings on the labor market, over better health, to higher economic growth. Besides the level of human capital, the degree of equality of educational opportunity is associated with positive outcomes like lower social inequality and higher intergenerational mobility of human capital and income. Compulsory schooling is crucial since it supplies people with the basic skills and sets the basis for further learning.¹ Therefore, student performance has gained increasing importance as an indicator for the level of human capital provided by a country's education system. Large scale international student performance studies like TIMSS, PIRLS and PISA have enabled researchers to draw on comparable data on educational achievement and have also spurred the attention devoted to the quality of education in compulsory schooling. Although the importance of human capital for economic growth and prosperity has been emphasized for a long time, internationally comparable measures of skills have intensified the debate on the quality of schooling systems greatly. Other measures such as educational credentials lack the comparability across educational systems. International studies on student performance instead apply the same test to all students and provide a constant measure across countries.

Most research in the field of educational production so far has concentrated on the benefits of resources for students, showing an all but clear-cut link between resources and student achievement. Recent studies have increasingly focused on the effects of institutional features such as central exams or school autonomy and have consistently found effects on student performance. This study contributes to the literature firstly by comparing the process of educational production between two countries and secondly by examining the relationship between schooling institutions and educational opportunities across a sample of countries in detail. Both a low level of educational achievement of students and a low degree of equality of educational opportunity are undesirable outcomes of a schooling system because great inequality in the acquisition of basic skills leads to even greater social and economic inequalities later on. Moreover, low educational opportunities can lead to a sub-optimal investment

1. Skills in general comprise both cognitive and non-cognitive abilities that determine the capability to accumulate knowledge and build up human capital and are measured (imperfectly) by educational achievement such as student performance.

in human capital and hence to a lower growth path. Therefore, the major questions that were addressed in this study is what factors determine educational achievement and in how far schooling institutions can be linked to educational opportunities. Knowing the determinants and the interplay between the latter two factors allows policy makers to influence educational achievement for different groups of students through the institutional framework of the schooling system.

This final chapter sums up the analysis of educational production conducted in this thesis. First, it reviews the analysis and summarizes the main findings of the preceding chapters. Second, it generates a synthesis of the findings and finally draws conclusions from the analysis.

5.1 Summary of findings

The review of the empirical literature revealed that student characteristics and family background have by far the greatest impact of all inputs into the production of student achievement. In contrast, the evidence on the effect of resources and school quality in general is less conclusive and fails to establish a consistent link between these kinds of inputs and educational outcomes. Research on the effects of institutions increasingly gains importance in the economics of education and underlines that certain institutions, like central exams, improve the incentive system for students and teachers alike and thereby increase the efficacy in the production of skills. However, research on institutional effects focuses mainly on the direct relationship with student achievement and little evidence is provided on the role of institutions for educational opportunities. The literature shows not only the great importance of student background for producing skills but also the large differences in educational opportunities across countries. Therefore, the question arises by which factors these differences can be explained and in how far the institutional setting of the schooling systems plays a role.

The possible interplay between institutions and educational opportunities has been laid out subsequently in a theoretical model of educational production. Therein, skills are both produced at home and at school. For the home production of skills, student ability, student effort, student background and parents' commitment to their child's education are the main inputs. At school, resources and schooling institutions are additional input factors that determine the educational success of students together with student background factors. Since the model is based on a multiplicative Cobb-Douglas production function, all inputs interact in the production of skills. The model assumes homogeneous student ability and allows for a heterogeneous student background. It is hence set up to analyze the difference in educational achievement between students with different social backgrounds. The equilibrium is determined by a joint maximization of students and parents over student effort and parents' commitment, respectively.

The model is then applied to examine the impact of changes in certain institutional factors on differences in performance between students with high and low social background, i.e.

the degree of equality of educational opportunity. Both the degree of streaming and the size of the private school sector affect the degree of differentiation in the school system. A higher differentiation offers a greater choice to students. However, when students from a high social background face fewer restrictions and thereby have greater choice regarding which school to attend, streaming and a large private school sector decrease the degree of educational opportunity, as shown by the theoretical model. A similar reasoning applies to the effect of school autonomy. A higher degree of school autonomy delegates decision making to a decentralized level. Parents can influence decisions more easily when these are taken at a local level than at a centralized level. When parents of students with a higher social background also show a higher degree of commitment to their child's education, they are more likely to take initiative at their school to the benefit of their child. Therefore, students with a high social background profit more from higher school autonomy relative to students from a low social background, irrespective of the overall impact of school autonomy on student performance. The final institutional feature whose effect is examined by the theoretical model is the amount of instruction time. A higher amount of instruction time decreases the importance of the home part of the production of skills. Therefore, student background is a less important determinant of student achievement and educational opportunity increases when students spend more time in school together. The findings of the theoretical analysis provide insights into which effects of certain institutions on educational opportunities can be expected and in which way the effects take place.

The first part of the empirical analysis focuses on the gap in performance between Finnish and German students and the effects that contribute to the difference in achievement as measured by test scores. Finnish students perform much better than German students at the age of fifteen in PISA, especially at the lower end of the test score distribution. While the gap between the two countries amounts to 30 points for the highest performing decile of students, it reaches over 75 test score points for the lowest decile. Estimating educational production functions separately for both countries shows that student achievement is determined to a higher degree by observable characteristics in Germany than in Finland. Especially the impact of social background is higher in Germany, which has a streamed schooling system with three major school types. In Finland, instead, all students attend the same school type in lower secondary education. The decomposition of the test score gap reveals that differences in student background and resources can hardly explain any part of the difference in performance. On the contrary, German students are, on average, endowed with a higher social background than Finnish students. Only for the lowest deciles of the student distribution, student background is more favorable for Finnish than for German students and can explain a part of the score gap. However, the use of resources and the transformation of student characteristics into student performance are estimated to be more effective in Finland than in Germany and can hence explain a large part of the test score gap. The main difference between the educational processes in the two countries is the degree of equality of educational opportunity, which is much higher in Finland than in Germany. However, the two-country analysis uses only the variation that is observable within each country. Differences in institutional features that appear only between the countries such as streaming cannot be analyzed by this approach. Although the information that is provided on students and schools in the PISA

data is extremely rich, the analysis is limited by the cross-sectional nature of the data. The identification of e.g. resource effects needs strong assumptions on selection and prior inputs and these results are therefore interpreted only cautiously.

The cross-country analysis that constitutes the second empirical analysis in this thesis draws on differences in institutional features of the schooling systems to explain the large variation in educational opportunities that was already observed in the comparison of two countries in the first empirical analysis. In order to consistently estimate the link between institutions and educational opportunities, the schooling systems of a sample of mostly European countries are analyzed at two stages, the end of primary education and the end of lower secondary education. The changes in student background effects are interacted with changes in the institutional features, such as the number of school types, the size of the private school sector, the amount of instruction time and school autonomy between the two stages. Thereby, country-specific effects can be largely controlled for and the institutional effects can be identified. The findings are that a high differentiation of the schooling system, which is either due to streaming or to a large private school sector, decreases educational opportunities for students with a lower social background. The stronger the use of streaming, as indicated by the number of school types, the greater is the effect. In the observed range of private school sector sizes, a larger share of students attending private schools steadily increases the impact of social background on student performance. The amount of instruction time, on the other hand, positively affects the degree of equality of educational opportunity. A higher degree of school autonomy benefits students whose parents have a high commitment to their child's education more relative to students whose parents care less about educational achievement. Most of the hypotheses from the theoretical model therefore find empirical support. The link between the findings from the model and the empirical parts are discussed in more detail in the following section.

Further findings are shifts in the importance of student background factors for educational achievement over student age. The attitude of parents towards learning becomes more important while family characteristics such as parents' origin or education are relatively less decisive for determining student achievement at later ages than in primary school. Further research, using institutional variation within countries, such as by region or province or, using a larger set of countries with comparable data, has to provide more evidence on these issues. Moreover, the results rely on the comparability of two cross-sectional studies and the quantitative information that is available on institutional features of schooling systems.

5.2 Synthesis

The synthesis links the findings of the literature review, the theoretical and two empirical analyses in this thesis. It explores in how far the results reinforce or oppose one another and what additional information is to be gained from a simultaneous interpretation of the results.

The comparison of the Finnish and German schooling systems reinforces the main conclusions from the literature review. Student background is by far the most important determinant of educational achievement. However, the impact of resources and school characteristics is low. Neither the observed teacher characteristics nor institutional variables have a significant effect on student performance. The cross-sectional data are *not* informative with respect to the performance gains caused by teachers, since these might be due to teachers' unobservable characteristics. These could be determined by differences in the labor market for teachers, which are claimed to be strong. School types are an important indicator of educational achievement in Germany. Their estimated effect is no causal effect of streaming because it includes the effect of selection into school types as well. The decomposition of the test score gap between the two countries also underlines that the effect of student background and thus the degree of equality of educational opportunity can vary greatly between countries. The second empirical analysis shows that the degree of streaming in a schooling system is an important factor for explaining these large differences in educational opportunities. Moreover, the decomposition analysis, in the first empirical analysis, reveals that primarily students from a low social background perform poorly in Germany compared to Finland. Besides streaming, the treatment of weak students in a class and additional support that is provided to them might explain this but these factors are not observed. The strong change of the gap in test scores along the test score distribution also emphasizes how important it is to look at specific subgroups or to consider the whole distribution. This relates to the evidence in the literature review showing that program effects differ between disadvantaged students and the average student. However, the results of quantile regressions provide no evidence that the effects of inputs estimated by the educational production function differ along the residual test score distribution.

The cross-country analysis in Chapter 4 picks up the remaining questions from the comparison between Finland and Germany. Instead of using the variation in institutions within the countries, it explains the observed variation in educational opportunities by the institutional variation between countries. Thereby, this part of the empirical analysis draws on the research prospects of the first part and confirms the conclusions. Moreover, it allows testing the hypotheses on the relationship between schooling institutions and educational opportunities from the theoretical model empirically.

The hypotheses concerning the negative relationship between streaming and the share of students attending private schools on the one hand and the equality of educational opportunity on the other are strongly supported by the empirical evidence. The model also predicts a lower impact of student background resulting from more instruction time, which is found in the empirical analysis. The reasoning is that the more time students spend in school together and the lower the share of home production of education, the less student background determines student performance. The empirical evidence also supports the hypothesis of a positive relationship between school autonomy and the effect of parents' attitude on student performance. When parents want to take initiative, this is easier in a decentralized system of decision making. These are the hypotheses from the theoretical model that have empirical support. However, the empirical analysis provides no significant evidence on the relationship between

parents' attitude and the amount of instruction time. This should be attributed to the unsatisfactory measurement of parents' attitude, namely only at the school level. Moreover, the link between social background and the number of school types and instruction time could only be established either for the number of books at home or parents' education, not for both indicators of social background. These results suggest that parents' education and the number of books at home represent different dimensions of social background. However, the theoretical model does not distinguish between the two when making hypotheses on the interaction with the institutional features. Furthermore, other institutions like central exams that were consciously ignored in the analysis, might have an effect on educational opportunities as well, be it through the interaction with other institutional features. Empirical analyses, in particular when being limited by the observed variation across countries, always have to restrict their focus on a specific section of the entire process, however.

Although most of the empirical findings in the two empirical analyses reinforce the conclusions from the literature review and the predictions by the theoretical model, only the direction of the effects and not the mechanisms or the specific theoretical argumentation is supported by the empirical facts. These underlying processes can only be analyzed using structural models based on richer data with more detailed information on institutions and student behavior as pointed out in the research prospects. Furthermore, many claims on differences between countries that rest on teachers' education, the art of teaching, the reputation of education and performance and general differences in the education systems that are not quantifiable cannot be addressed in the empirical analyses.

5.3 Policy conclusions

This study provides both an in-depth theoretical and empirical examination of the role of schooling institutions in the production of human capital. The main contribution lies in the analysis of the interplay between institutions and educational opportunities. After having reviewed the main findings from the analysis and linked them in a synthesis, the study is finalized by drawing conclusions from the key findings. Moreover, future research prospects in this area are outlined.

The publication of the PISA 2000 results led to a great debate in countries that performed below the OECD average, such as Germany. In addition to the low average achievement of leading industrial countries, the large inequality in skills is an unpromising result of the student performance study. As a consequence, the German schooling system is compared to more successful schooling systems in PISA. In particular the Finnish system, that produces both high mean scores and a low spread of measured achievement in PISA, serves as a role model for European schooling systems. The analysis of the difference in student performance between Finland and Germany leads one to conclude that it is not so much the inputs that have to be adjusted in Germany but rather the schooling institutions, which set the framework in which education is produced in the schooling system. This could diminish the high

dependence of educational success on student background and improve the effectiveness in the use of resources.

The greatest challenge for the German schooling system is to promote the achievement of low performing students. This can be accomplished by increasing the educational opportunity for disadvantaged children from a low social background. A lower differentiation of the schooling system was shown to increase the degree of equality of educational opportunity. Both early streaming in lower secondary education and a large size of the private school sector contribute to a higher differentiation and should hence be reduced when the achievement of the socially disadvantaged children is to be promoted. There seems to be no adverse effect but rather a positive impact of a reduction of streaming on mean performance of students (cf. Hanushek and Woessmann, 2006; Maurin and McNally, 2006). Moreover, policy makers should be aware that greater school autonomy also increases the influence of parents. When parents with a high social background also show more interest in their child's education and take influence in their child's favor, school autonomy may increase the gap in educational opportunities between students from different social backgrounds. On the other hand, more instruction time decreases the impact of family background on performance. In particular schooling systems with half-day schools, in which students do not stay over lunch, have a means to improve the achievement of low performers, who are mostly socially disadvantaged, by extending the instruction and care time. While the former policies of abandoning or delaying streaming and restricting the establishment of private schools and school autonomy is mostly cost-neutral, at least in the longer run, the increase in instruction and care time requires additional funding of the schooling system.

It should be noted that the actual implementation of institutions in a specific schooling system cannot be evaluated and was not intended to be the focus of this study. When private schools are not more selective than public schools, for example, or when the allocation to school types is solely based on student ability and not on background and all school types succeed in producing skills, these institutions should not affect the degree of equality of educational opportunity. However, the comparison of the schooling systems in the considered sample led to the relationship between institutions and educational opportunities and the policy conclusions described above, bearing in mind the limitations of the data.

5.4 Research prospects

The analyses conducted in this study expose remaining open questions and identify the scope for further research. A great field of research that is still to be explored is the underlying process by which institutions affect educational production. This process is specific to each schooling system and can only be examined with a profound knowledge of the institutional system. Examples are the process of allocation of students to different school types in a streamed system and the resulting peer effects within each school type. Although evidence on peer effects is already provided by the literature, in particular for U.S. schools, the interplay with further institutional features like streaming or with the use of resources is hardly exam-

ined. An example would be the interplay between the effect of class size and peer effects, as suggested by Lazear (2001). Furthermore, the interaction between different institutions and their effect on both the level and the inequality of educational achievement remain to be investigated. In order to explain the overall difference in educational opportunities it would be of interest to consider factors like tradition and culture as well. These are not the focus of this study and are largely eliminated by the empirical strategy. When the interest is on the absolute level of educational opportunities, e.g. to explain differences in wage inequality between countries, country-specific effects should not be neglected.

This study made a contribution by considering several institutional features simultaneously but did not allow for an explicit interaction between institutions. The available data clearly limit the possibility of identifying the various effects. Longitudinal cross-country data would allow exploiting both the variation across countries and over time on the level of the student and yield further insights. The use of panel data would also allow for an improved identification of causal effects. Therefore, it would be most valuable to construct repeated student performance studies such as PISA or PIRLS as panel data instead of cross-sectional studies. One alternative for most countries is to further study the comparability of various cross-sectional studies and exploit their information, as has been done in this study as well. The link between outcomes in student performance studies and labor market outcomes is an additional field of research for which empirical evidence is restricted to few countries so far. Providing stronger evidence on the relationship between distributions in student achievement and wages or employment prospects would help to shed light on the importance of the allocation of society's resources.

Literature

- Adams, Ryan and Margaret Wu (2002), PISA 2000 technical report, OECD, Paris.
- Aghion, Philippe and Peter Howitt (1998), *Endogenous growth theory*, MIT Press, Cambridge.
- Akerhielm, Karen (1995), Does class size matter?, *Economics of Education Review* 14(3), 229-241.
- Ammermueller, Andreas (2005), Educational opportunities and the role of institutions, ZEW Discussion Paper No. 05-44, Mannheim.
- Ammermueller, Andreas (2006), PISA: What makes the difference? Explaining the gap in PISA test scores between Finland and Germany, *Empirical Economics*, forthcoming.
- Ammermueller, Andreas (2007), Poor background or low returns? Why immigrant students in Germany perform so poorly in PISA, *Education Economics*, forthcoming.
- Ammermueller, Andreas and Peter Dolton (2006), Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA, ZEW Discussion Paper No. 06-60, Mannheim.
- Ammermueller, Andreas, Hans Heijke and Ludger Woessmann (2005), Schooling quality in Eastern Europe: Educational production during transition, *Economics of Education Review* 24, 579-599.
- Ammermueller, Andreas and Joern-Steffen Pischke (2006), Peer effects in European primary schools: Evidence from PIRLS, NBER Working Paper No. 12180, Cambridge.
- Ammermueller, Andreas and Andrea Maria Weber (2005), Educational attainment and returns to education in Germany, ZEW Discussion Paper No. 05-17, Mannheim.
- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King and Michael Kremer (2002), Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment, *American Economic Review* 92(5), 1535-1558.
- Angrist, Joshua, Eric Bettinger and Michael Kremer (2004), Long-term consequences of secondary school vouchers: Evidence from administrative records in Colombia, NBER Working Papers No. 10713, Cambridge.
- Angrist, Joshua and Jonathan Guryan (2005), Does teacher testing raise teacher quality? Evidence from state certification requirements, IZA Discussion Paper No. 1500, Bonn.
- Angrist, Joshua and Kevin Lang (2002), How important are classroom peer effects? Evidence from Boston's Metco Program, NBER Working Papers No. 9263, Cambridge.
- Angrist, Joshua and Victor Lavy (1999), Using Maimonides' rule to estimate the effect of class size on scholastic achievement, *Quarterly Journal of Economics* 114(2), 533-575.

- Angrist, Joshua and Victor Lavy (2002), New evidence on classroom computers and pupil learning, *Economic Journal* 112, 735-765.
- Anweiler, Oskar, Ursula Boos-Nünning, Günter Brinkmann, Detlef Glowka, Dieter Goetze, Wolfgang Hörner, Friedrich Kuebart and Hans P. Schäfer (1996), *Bildungssysteme in Europa*, Beltz, Weinheim.
- Artelt, Cordula, Ulrich Schiefele and Wolfgang Schneider (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, 16(3), 363-383.
- Asplund, Rita and Erling Barth (eds) (2005), *Education and wage inequality in Europe, A literature review*, ETLA, Sarja B 212 Series, Helsinki.
- Barro, Robert (1999), Human capital and growth in cross-country regressions, *Swedish Economic Policy Review*, 6(2), 237-277.
- Barro, Robert (2001), Human capital and growth, *American Economic Review*, Papers and Proceedings 91(2), 12-17.
- Barro, Robert and Jong Wha Lee (1996), International measures of schooling years and schooling quality, *American Economic Review*, Papers and Proceedings 86(2), 218-23.
- Baumert, Jürgen and Deutsches PISA-Konsortium (eds) (2001), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*, Waxmann, Münster.
- Baumert, Jürgen and Deutsches PISA-Konsortium (eds) (2003), *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Waxmann, Münster.
- Becker, Gary (1964/1993), *Human capital: A theoretical and empirical analysis, with special reference to education*, Third Edition, University of Chicago Press, Chicago and London.
- Becker, Gary (1967), *Human capital and personal distribution of income: An analytical approach*, Ann Arbor, University of Michigan Press.
- Bedard, Kelly and Christopher Ferrall (2003), Wage and test score dispersion: Some international evidence, *Economics of Education Review* 22(1), 31-43.
- Benabou, Roland (1996), Equity and efficiency in human capital investment: The local connection, *Review of Economic Studies* 63, 237-264.
- Benabou, Roland (2000), Unequal societies: Income distribution and the social contract, *American Economic Review* 90, 96-129.
- Ben-Porath, Yoram (1967), The production of human capital and the life cycle of earnings, *Journal of Political Economy* 75(4), 352-365.
- Berger, Mark and Paul Leigh (1989), Schooling, self-selection, and health, *Journal of Human Resources* 24(3), 433-55.
- Betts, Julian (2001), The impact of school resources on women's earnings and educational attainment: Findings from the National Longitudinal Survey of Young Women, *Journal of Labour Economics* 19(3), 635-657.
- Bishop, John (1997), The effect of national standards and curriculum-based exams on achievement, *American Economic Review*, Papers and Proceedings 87(2), 260-264.

- Bishop, John (1999), Are national exit examinations important for educational efficiency?, *Swedish Economic Policy Review* 6(2), 349-401.
- Bishop, John and Ferran Mane (2005), Raising academic standards and vocational concentrators: Are they better off or worse off?, *Education Economics* 13(2), 171-187.
- Bishop, John and Ludger Woessmann, (2004), Institutional effects in a simple model of educational production, *Education Economics* 12(1), 17-38.
- Björklund, Anders and Mats Jäntti (1997), Intergenerational income mobility in Sweden compared to the United States, *American Economic Review* 87(5), 1009-1018.
- Black, Sandra, Paul Devereux and Kjell Salvanes (2005), Why the apple doesn't fall far: Understanding the intergenerational transmission of education, *American Economic Review* 95, 437-449.
- Blanden, Jo, Paul Gregg and Stephen Machin (2005), Educational inequality and intergenerational mobility, in: S. Machin and A. Vignoles (eds), *What's the good of education?*, Princeton University Press, Princeton.
- Blau, Francine and Lawrence Kahn (1996), International differences in male wage inequality: Institutions versus market forces, *Journal of Political Economy* 104(4), 791-837.
- Blau, Francine and Lawrence Kahn (2005), Do cognitive test scores explain higher U.S. wage inequality?, *Review of Economics and Statistics* 87(1), 184-193.
- Blinder, Alan (1973), Wage discrimination: Reduced form and structural estimates, *Journal of Human Resources* 8(4), 436-455.
- BLK (2003), BLK-Bildungsfinanzbericht 2001/2002, Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, Heft 108-II, Bonn.
- Borghans, Lex (2002), Class teaching and individual instruction, Maastricht University, mimeo.
- Borghans, Lex, Francis Green and Ken Mayhew (2001), Skills measurement and economic analysis: An introduction, *Oxford Economic Papers* 3, 375-384.
- Borghans, Lex and Hans Heijke (2005), The production and use of human capital: Introduction, *Education Economics* 13(2), 133-142.
- Borghans, Lex, Huub Meijers and Bas ter Weel (2007), The role of noncognitive skills in explaining cognitive test scores, *Economic Inquiry* 45, forthcoming.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther and Renate Valtin (eds) (2003), *Erste Ergebnisse aus IGLU, Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*, Waxmann, Münster.
- Braddock, Jomills (1990), Tracking: Implications for student race-ethnic subgroups, Technical Report No. 1, Center for Research on Effective Schooling for Disadvantaged Students, Baltimore.
- Bressoux, Pascal, Francis Kramarz and Corinne Prost (2005), Teachers' training, class size and students' outcomes: Evidence from third grade classes in France, mimeo.
- Brown, Giorgina, John Micklewright, Sylke Schnepf and Robert Waldmann (2005), Cross-national surveys of learning achievement: How robust are the findings?, IZA Discussion Paper No. 1652, Bonn.

- Brunello, Giorgio and Daniele Checchi (2005), School quality and family background in Italy, *Economics of Education Review*, forthcoming.
- Brunello, Giorgio and Daniele Checchi (2006), Does school quality affect educational opportunity? New international evidence, IZA Discussion Paper No. 2348, Bonn.
- Brunello, Giorgio, Massimo Giannini and Kenn Ariga (2004), The optimal timing of school tracking, IZA Discussion Paper No. 995, Bonn.
- Brunello, Giorgio and Lorenzo Rocco (2004), Diploma no problem: Can private schools be of lower quality than public schools?, IZA Discussion Paper No. 1336, Bonn.
- Cappellari, Lorenzo (2004), High school types, academic performance and early labour market outcomes, IZA Discussion Paper No. 1048, Bonn.
- Card, David (1999), The causal effect of education on earnings, in: O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3a, North Holland, Amsterdam.
- Card, David and Alan B. Krueger (1992), Does school quality matter? Returns to education and the characteristics of public schools in the United States, *Journal of Political Economy* 100(1), 1-40.
- Card, David and Alan B. Krueger (1994), Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review* 84, 772-793.
- Card, David and Alan B. Krueger (1996), Labor market effects of school quality: Theory and evidence, in: G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success*, Washington D.C., Brookings Institution.
- Carneiro, Pedro, James Heckman and Edward Vytlacil (2003), Understanding what instrumental variables estimate: Estimating marginal and average returns to education, http://www.stanford.edu/~vytlacil/_papers/school_all_7_21_03_gko.pdf.
- Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederick Weinfeld and Robert York (1966), *Equality of educational opportunity: Summary report*, U.S. Government Printing Office, Washington, D.C.
- Cunha, Flavio, James Heckman, Lance Lochner and Dimitriy Masterov (2005), Interpreting the evidence on life-cycle skill formation, NBER Working Paper No. 11331, Cambridge.
- Dearden, Lorraine, Javier Ferri and Costas Meghir (2002), The effect of school quality on educational attainment and wages, *Review of Economics and Statistics* 84(1), 1-20.
- Dearden, Lorraine, Stephen Machin and Howard Reed (1997), Intergenerational mobility in Britain, *Economic Journal* 107(1), 47-66.
- Deaton, Angus (1997), *The analysis of household surveys*, John Hopkins University Press, Baltimore and London.
- Dee, Thomas (2005), A teacher like me: Does race, ethnicity, or gender matter?, *American Economic Review* 95(2), 158-65.
- De Gregorio, Jose (1996), Borrowing constraints, human capital accumulation, and growth, *Journal of Monetary Economics* 37, 49-71.

- De Leeuw, Jan and Ita Kreft (1995), Questioning multilevel models, *Journal of Educational Behavioural Statistics* 20, 171-189.
- Devroye, Dan and Richard Freeman (2001), Does inequality in skills explain inequality of earnings across advanced countries?, NBER Working paper No. 8140, Cambridge.
- Dolton, Peter (2004), Recruiting high-quality teachers, *CESifo DICE Report* 2(4), 11-20.
- Dolton, Peter, Rosalind Levacic and Anna Vignoles (2004), The economic impact of schooling resources, in: C. Sofer (ed.), *Human capital over the life cycle*, Edward Elgar, 36-52.
- Dolton, Peter and Anna Vignoles (1999), The impact of school quality on labour market success in the UK, University of Newcastle Discussion Paper 98-03, Newcastle.
- Dolton, Peter and Anna Vignoles (2000), The effects of school quality on pupil outcome: An overview, in: H. Heijke and J. Muysken (eds), *Education and training in a knowledge-based economy*, Macmillan, London.
- Dolton, Peter, Gerald Makepeace and Oscar Marcenaro-Gutierrez (2005), Career progression: Getting-on, getting-by and going nowhere, *Education Economics* 13(2), 237-255.
- Donald, Stephen and Kevin Lang (2004), Inference with difference in differences and other panel data, Working Paper, University of Texas.
- Drucker, Peter (1969), *The age of discontinuity; Guidelines to our changing society*, Harper and Row, New York.
- Dustmann, Cristian, Najma Rajah and Arthur Soest (2002), Class size, education, and wages, *Economic Journal* 113(485), F99-F120.
- Economist (2005), Sorting out schools, *Economist* 377 (8457), 54-54.
- Economist (2006a), Danger: Good schools ahead, *Economist* 378 (8462), 55-55.
- Economist (2006b), Wasting brains, *Economist* 378 (8464), Special Section 6-8.
- Epple, Dennis, Elizabeth Newlon and Richard Romano (2002), Ability tracking, school competition, and the distribution of educational benefits, *Journal of Public Economics* 83, 1-48.
- Epple, Dennis and Richard Romano (1998), Competition between private and public schools, vouchers, and peer-group effects, *American Economic Review* 88(1), 33-62.
- Eurybase (2003), The information network on education in Europe, <http://www.eurydice.org>, visited October 24th 2003.
- Feldhusen, John (1989), Synthesis of research on gifted youth, *Educational Leadership* 46(6), 6-11.
- Feldman, Jacob, Diane Makuc, Joel Kleinman and Joan Cornoni-Huntley (1989), National trends in educational differentials in mortality, *American Journal of Epidemiology* 129(5), 919-933.
- Fertig, Michael (2003), Who's to blame? The determinants of German students' achievement in the PISA 2000 study, RWI Discussion Paper No. 4, Essen.
- Fertig, Michael and Jochen Kluve (2005), The effect of age at school entry on educational attainment in Germany, RWI Discussion Paper No. 27, Essen.

- Figlio, David (1999), Functional form and the estimated effects of school resources, *Economics of Education Review* 18, 241-252.
- Figlio, David (2005), Names, expectations and the black-white test score gap, NBER Working Paper No. 11195, Cambridge.
- Figlio, David and Jens Ludwig (2000), Sex, drugs, and catholic schools: Private schooling and non-market adolescent behaviors, NBER Working Paper 7990, Cambridge.
- Figlio, David and Marianne Page (2000), School choice and the distributional effects of ability tracking: Does separation increase equality?, NBER Working Paper No. 8055, Cambridge.
- Fredriksson, Peter and Björn Öckert (2004), Is early learning really more productive? The effect of school starting age on school and labor market performance, mimeo, Uppsala University.
- Fuchs, Thomas and Ludger Woessmann (2004), Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school, CESifo Working Paper No. 1321, Munich.
- Fuchs, Thomas and Ludger Woessmann (2006), What accounts for international differences in student performance? A re-examination using PISA data, *Empirical Economics*, forthcoming.
- Galindo-Rueda, Fernando and Anna Vignoles (2004), The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability, IZA Discussion Paper No. 1245, Bonn.
- Galor, Oded and Joseph Zeira (1993), Income distribution and macroeconomics, *Review of Economic Studies* 60(1), 35-52.
- Gamoran, Adam (1989), Measuring curriculum differentiation, *American Journal of Education* 97, 129-143.
- Garcia, Jaume, Pedro Hernández and Angel López-Nicolás (2001), How wide is the gap? An investigation of gender wage differences using quantile regression, *Empirical Economics* 26, 149 – 167.
- Gautschi, Eliane (2005), What we can learn from Finland: Facts and reflections on the PISA study, *Current concerns* 4.
- Gibbons, Steve (2002), Geography, resources and primary school performance, CEE Discussion Papers No. 25, London.
- Glewwe, Paul (1997), Estimating the impact of peer group effects on socio-economic outcomes, *Economics of Education Review* 16(1), 39-43.
- Glewwe, Paul (1999), Why does mother's schooling raise child health in developing countries? Evidence from Morocco, *Journal of Human Resources* 34(1), 124-159.
- Goldhaber, Dan and Dominic Brewer (1997), Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity, *Journal of Human Resources* 32(3), 505-523.
- Goldhaber, Dan, Dominic Brewer and Deborah Anderson (1999), A three way error components analysis of education productivity, *Education Economics* 7(3), 199-208.

- Goldstein, Harvey (1995), *Multilevel statistical models*, Arnold, London.
- Goldstein, Rich (1996), Patterns of missing data, *STATA Technical Bulletin* 32, 12-13.
- Graddy, Kathryn and Margaret Stevens (2003), The impact of school inputs on student performance: An empirical study of private schools in the United Kingdom, CEPR Discussion Paper Series No. 3776, London.
- Graham, Bryan (2004), Identifying social interactions through excess variance contrasts, Harvard University, mimeo.
- Green, Andy, Alison Wolf and Tom Leney (1999), *Convergence and divergence in European education and training systems*, University of London, London.
- Green, David and Craig Riddell (2003), Literacy and earnings: An investigation of the interaction of cognitive and unobserved skills in earnings generation, *Labour Economics* 10, 165-184.
- Groot, Wim and Hessel Oosterbeek (1994), Earnings effects of different components of schooling; Human capital versus screening, *Review of Economics and Statistics* 76, 317-321.
- Gundlach, Erich, Desmond Rudman and Ludger Woßmann (2002). Second thoughts on development accounting, *Applied Economics* 34(11), 1359-1369.
- Gundlach, Erich and Ludger Woessmann (2004), Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz, in: U. Backes-Gellner and P. Moog (eds), *Ökonomie der Evaluation von Schulen und Hochschulen*, Duncker und Humboldt, Berlin, 15-52.
- Gundlach, Erich, Ludger Woessmann and Jens Gmelin (2001), The decline of schooling productivity in OECD countries, *Economic Journal* 111(471), C135-C147.
- Haegeland, Torbjörn, Oddbjørn Raaum and Kjell Salvanes (2005), Pupil achievement, school resources and family background, IZA Discussion Paper No. 1459, Bonn.
- Häkkinen, Lilda, Tanja Kirjavainen and Roope Uusitalo (2003), School resources and student achievement revisited: New evidence from panel data, *Economics of Education Review* 22, 329-335.
- Hall, Robert and Charles Jones (1999). Why do some countries produce so much more output per worker than others?, *Quarterly Journal of Economics* 114(1), 83-116.
- Hambleton, Ronald and Hariharan Swaminathan (1989), *Item response theory. Principles and applications*, Kluwer-Nijhof, Boston.
- Hanushek, Eric (1986a), The economics of schooling: Production and efficiency in public schools, *Journal of Economic Literature* 24(3), 1141-1177.
- Hanushek, Eric (1986b), School resources and student performance, in: G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success*, Washington D.C., Brookings Institution.
- Hanushek, Eric (1989), The impact of differential expenditures on school performance, *Educational Researcher* 18(4), 45-51.
- Hanushek, Eric (2003), The failure of input-based schooling policies, *Economic Journal* 113(485), F64-F98.

- Hanushek, Eric, John Kain, Jacob Markman and Steven Rivkin (2003), Does peer ability affect student achievement?, *Journal of Applied Econometrics* 18(5), 527-544.
- Hanushek, Eric and Dennis Kimko (2000), Schooling, labor-force quality and the growth of nations, *American Economic Review* 90(5), 1184-1208.
- Hanushek, Eric and Javier Luque (2003), Efficiency and equity in schools around the world, *Economics of Education Review* 22(5), 481-502.
- Hanushek, Eric and Steven Rivkin (2003), How to improve the supply of high quality teachers, Paper Prepared for the Brookings Papers on Educational Policy May 21-22 2003.
- Hanushek, Eric, Steven Rivkin and Lori Taylor (1996), Aggregation and the estimated effects of school resources, *Review of Economics and Statistics*, 78(4), 611-627.
- Hanushek, Eric and Ludger Woessmann (2006), Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries, *Economic Journal* 116 (510), 2006, C63-C76.
- Haveman, Robert and Barbara Wolfe (2000), Accounting for the social and non-market benefits of education, in: J. Helliwell (ed.), *The contribution of human and social capital to sustained economic growth and well-being: International Symposium Report*, OECD and HRDC, 221-250.
- Heckman, James (2000), Policies to foster human capital, *Research in Economics* 54(1), 3-56.
- Heckman, James (2005), Lessons from the technology of skill formation, NBER Working Paper No. 11142, Cambridge.
- Hedges, Larry V., Richard D. Laine and Rob Greenwald (1994), Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes, *Educational Researcher* 23(3), 5-14.
- Heijke, Hans and Joan Muysken (2000), Introduction, in: H. Heijke and J. Muysken (eds), *Education and training in a knowledge-based economy*, Macmillan, London.
- Hidalgo-Hidalgo, Marisa (2005), Peer group effects and optimal education system, WP-AD 2005-12, Instituto Valenciano de Investigaciones Economicas, S.A., Valencia.
- Hoxby, Caroline (1999), The productivity of schools and other local public goods providers, NBER Working Paper No. 6911, Cambridge.
- Hoxby, Caroline (2000), The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics* 115(4), 1239-1285.
- Hoxby, Caroline (2003), Peer effects in the classroom: Learning from gender and race variation, NBER Working Paper No. 7867, Cambridge.
- Jann, Ben (2005), Standard errors for the Blinder–Oaxaca decomposition, http://repec.org/dsug2005/oaxaca_se_handout.pdf.
- Jürges, Hendrik, Kerstin Schneider and Felix Büchel (2005), The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany, *Journal of the European Economic Association* 3, 1134-1155.

- Jürges, Hendrik, Kerstin Schneider and Wolfram Richter (2004), Teacher quality and incentives, Theoretical and empirical effects of standards on teacher quality, mimeo, MEA, Mannheim.
- Juhn, Chinhui, Kevin Murphy and Brooks Pierce (1993), Wage inequality and the rise in the returns to skills, *Journal of Political Economy* 101(3), 410-442.
- Klös, Hans-Peter and Reinhold Weiß (2003), *Bildungsbenchmarking Deutschland*, Köln.
- KMK (2004), Pressemitteilung der Kultusministerkonferenz, Bonn.
- Krueger, Alan B. (1999), Experimental estimates of education production functions, *Quarterly Journal of Economics* 114(2), 497-532.
- Krueger, Alan B. (2003), Economic considerations and class size, *Economic Journal* 113, F34-F63.
- Krueger, Alan B. and Mikael Lindahl (2001), Education for growth: Why and for whom?, *Journal of Economic Literature* 39, 1101-1136.
- Krueger, Alan B. and Diane Whitmore (1999), The effects of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR, *Economic Journal* 111(468), 1-28.
- Lauer, Charlotte (2000), Gender wage gap in West Germany: How far do gender differences in human capital matter?, ZEW Discussion Paper No. 00-07, Mannheim.
- Lavy, Victor (2003), Paying for performance: The effect of teachers' financial incentives on students' scholastic outcomes, CEPR Discussion Papers No. 3862, London.
- Lavy, Victor (2004), Performance pay and teachers' effort, productivity and grading efforts, NBER Working Paper No. 10622, Cambridge.
- Lazear, Edward (2001), Educational production, *Quarterly Journal of Economics* 113(3), 777-803.
- Lazear, Edward (2003), Teacher incentives, *Swedish Economic Policy Review* 10, 179-214.
- Lee, Jong-Wha and Robert Barro (1997), Schooling quality in a cross section of countries, *Economica* 68(127), 465-88.
- Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek and Dinand Webbink (2004), The effect of extra funding for disadvantaged pupils on achievement, IZA Discussion Paper No. 1122, Bonn.
- Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek and Dinand Webbink (2006), Expanding schooling opportunities for 4-year-olds, IZA Discussion Paper No. 2434, Bonn.
- Leuven, Edwin, Hessel Oosterbeek and Hans van Ophem (2005), Explaining international differences in male wage inequality by differences in demand and supply of skill, *Economic Journal*, forthcoming.
- Levacic, Rosalind and Anna Vignoles (2002), Researching the links between school resources and student outcomes in the UK: A review of issues and evidence, *Education Economics* 10(3), 313-331.

- Lochner, Lance and Enrico Moretti (2004), The effect of education on crime: Evidence from prison-inmates, arrests and self-reports, *American Economic Review* 94(1), 155-189.
- Machin, Stephen and Sandra McNally (2004), The literacy hour, IZA Discussion Paper No. 1005, Bonn.
- Machin, Stephen, Sandra McNally and Costas Meghir (2004), Improving pupil performance in English secondary schools: Excellence in Cities, *Journal of the European Economic Association* 2(2-3), 396-405.
- Mankiw, Gregory, David Romer and David Weil (1992), A contribution to the empirics of economic growth, *Quarterly Journal of Economics* 107(2), 408-437.
- Manski, Charles (1993), Identification of endogenous social effects: The reflection problem, *Review of Economic Studies* 60(3), 531-542.
- Marshall, Alfred (1890/1922), *Principles of Economics: An Introductory Volume*, Eighth Edition, Macmillan, London.
- Maurin, Eric and Sandra McNally (2006), Selective schooling, London School of Economics, mimeo.
- McIntosh, Steven and Anna Vignoles (2001), Measuring and assessing the impact of basic skills on labour market outcomes, *Oxford Economic Papers* 53, 453-481.
- McMahon, Walter (1999), *Education and development: Measuring the social benefits*, Oxford University Press, Oxford.
- Meghir, Costas and Marten Palme (2005), Educational reform, ability and family background, *American Economic Review* 95(1), 414-424.
- Mejia, Daniel and Marc St-Pierre (2005), Unequal opportunities and human capital formation, CESifo Working Paper No. 1383, Munich.
- Meyer, Bruce (1995), Natural and quasi-experiments in economics, *Journal of Business & Economic Statistics* 13, 151-161.
- Mincer, Jacob (1974), *Schooling, experience, and earnings*, National Bureau of Economic Research, New York.
- Mislevy, Robert (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika* 56, 177-196.
- Moulton, Brent (1986), Random group effects and the precision of regression estimates, *Journal of Econometrics* 32(3), 385-397.
- Moulton, Brent (1990), An illustration of a pitfall in estimating the effects of aggregate variables on micro units, *Review of Economics and Statistics* 72(2), 334-338.
- Mukherjee, Chandan, Howard White and Marc Wuyts (1998), *Econometrics and data analysis for developing countries*, Routledge, London and New York.
- Mulligan, Casey (1999), Galton versus the human capital approach to inheritance, *Journal of Political Economy* 107(6), 184-224.
- Mullis, Ina, Michael O. Martin, Eugene J. Gonzalez and Ann M. Kennedy, (2003), *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*, Chestnut Hill, Boston College.

- Murnane, Richard, John Willett, Jay Braatz and Yves Duhaldeborde (2001), Do different dimensions of male high school students' skills predict labour market success a decade later? Evidence from the NLSY, *Economics of Education Review* 20(4), 311-320.
- Murnane, Richard, John Willett, Yves Duhaldeborde and John Tyler (2000), How important are the cognitive skills of teenagers in predicting subsequent earnings?, *Journal of Policy Analysis and Management* 19(4), 547-568.
- Murnane, Richard, John Willett and Frank Levy (1995), The growing importance of cognitive skills in wage determination, *Review of Economics and Statistics* 77(2), 251-66.
- Neal, Derek (1997), The Effects of Catholic Secondary Schooling on Educational Achievement, *Journal of Labor Economics* 15(1), 98-123.
- Neal, Derek and William Johnson (1996), The role of pre-market factors in black-white wage differences, *Journal of Political Economy* 104, 869-895.
- Nechyba, Thomas (2000), Mobility, targeting, and private-school vouchers, *American Economic Review* 90, 130-146.
- Nickell, Stephen (2004), Poverty and worklessness in Britain, *Economic Journal* 114, C1-C25.
- Oakes, Jeannie (1985), *Keeping track: How schools structure inequality*, Yale University Press, New Haven.
- Oaxaca, Ronald (1973), Male-female wage differentials in urban labor markets, *International Economic Review* 14, 693-709.
- OECD (2001), *Knowledge and skills for life: First results from PISA 2000*, OECD, Paris.
- OECD (2002), *Education at glance 2002*, OECD, Paris.
- OECD (2004), *Learning for tomorrow's world – First results from PISA 2003*, OECD, Paris.
- Parsons, Samantha and John Bynner (1997), Numeracy and employment, *Education and Training* 39, 43-51.
- Pekkarinen, Tuomas, Roope Uusitalo and Sari Pekkala (2006), Education policy and intergenerational income mobility: Evidence from the Finnish comprehensive school reform, IZA Discussion Paper No. 2204, Bonn.
- Pischke, Joern-Steffen and Alan Manning (2006), Comprehensive versus selective schooling in England and Wales: What do we know?, NBER Working Paper No. 12176, Cambridge.
- Prescott, Edward (1998), Needed: A theory on total factor productivity, *International Economic Review*, 39(3), 525-552.
- Puhani, Patrick and Andrea M. Weber (2005), Does the early bird catch the worm? Instrumental variable estimates of educational effects of age of school entry in Germany, IZA Discussion Paper No. 1827, Bonn.
- Rivkin, Steven, Eric Hanushek and John Kain (2005), Teachers, schools, and academic achievement, *Econometrica* 73(2), 417-458.

- Rangvid, Beatrice (2003), Educational peer effects: Quantile regression evidence from Denmark with PISA 2000 data, Kapitel 3 i "Do schools matter?", phd-afhandling nr. 2003:9, Aarhus School of Business.
- Robertson, Donald and James Symons (1996), Do peer groups matter? Peer group versus schooling effects on academic attainment, Discussion Paper No. 311, Centre for Economic Performance, LSE, London.
- Robin, Stephane and Maresa Sprietsma (2003), Characteristics of teaching institutions and students' performance: New empirical evidence from OECD data, IRES Discussion Paper No. 28, Louvain-la-Neuve.
- Roemer, John (1998), *Equality of opportunity*, Harvard University Press, Cambridge.
- Romer, Paul (1990), Endogenous technological change, *Journal of Political Economy* 98(5), 71-102.
- Rosenbaum, James (1976), *Making inequality: The hidden curriculum high school tracking*, Wiley, New York.
- Rouse, Cecilia (1998), Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program, *Quarterly Journal of Economics* 113(2), 553-602.
- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2004), *Erfolge im Ausland – Herausforderungen im Inland, Jahresgutachten 2004/05*, Metzler-Poeschel, Stuttgart.
- Schafer, Joseph and Nathalie Schenker (2000), Inference with imputed conditional means, *Journal of the American Statistical Association* 95, 144-154.
- Schneeweis, Nicole and Rudolf Winter-Ebmer (2005), Peer effects in Austrian schools, Department of Economics, University of Linz, Working Paper No. 0502.
- Schuetz, Gabriela, Heinrich Ursprung and Ludger Woessmann (2005), Education policy and equality of opportunity, CESifo Working Paper No. 1518.
- Schultz, Theodore W. (1961), Investment in human capital, *American Economic Review*, 51(1), 1-17.
- Silverman, Bernard (1986), *Density estimation for statistics and data analysis*, Chapman & Hall, London.
- Slavin, Robert E. (1990), Achievement effects of ability grouping in secondary schools: A best-evidence synthesis, *Review of Educational Research* 60(3), 471-499.
- Smith, Adam (1776/1976), *An inquiry into the nature and causes of the wealth of nations*, in: R. Campbell and A. Skinner (eds), Glasgow Edition, Clarendon Press, Oxford.
- Spiess, C. Katharina, Felix Büchel and Gerd Wagner (2003), Children's school placement in Germany: Does kindergarten attendance matter?, IZA Discussion Paper No. 722, Bonn.
- Tinbergen, Jan (1975), *Income distribution, analysis and policies*, North-Holland, Amsterdam and Oxford.
- Todd, Petra and Kenneth Wolpin (2003), On the specification and estimation of the production function for cognitive achievement, *Economic Journal* 113(485), F3-F33.

- Todd, Petra and Kenneth Wolpin (2004), The production of cognitive achievement in children: Home, school and racial test score gaps, PIER Working Paper 04-019, Philadelphia.
- Väljjarvi, Jouni, Pirjo Linnakylä, Pekka Kupari, Pasi Reinikainen and Inga Arffman (2002), The Finnish success in PISA – and some reasons behind it, <http://www.jyu.fi/ktl/pisa/publication1.pdf>.
- Vandenberghe, Vincent and Stéphane Robin (2004), Evaluating the effectiveness of private education across countries: A comparison of methods, *Labour Economics* 11(4), 487-506.
- Vila, Luis (2000), The non-monetary benefits of education, *European Journal of Education*, 35, 21-32.
- Waldinger, Fabian (2006), Does tracking affect the importance of family background on students' test scores?, London School of Economics, mimeo.
- White, Halbert (1980), A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity, *Econometrica* 48(4), 817-828.
- Woessmann, Ludger (2002), *Schooling and the quality of human capital*, Kieler Studien 319, Springer, Berlin.
- Woessmann, Ludger (2003), Schooling resources, educational institutions, and student performance: The international evidence, *Oxford Bulletin of Economics and Statistics* 65(2), 117-170.
- Woessmann, Ludger (2004), How equal are educational opportunities? Family background and student achievement in Europe and the United States, CESifo Working Paper No. 1162, Munich.
- Woessmann, Ludger (2005), The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA, *Education Economics* 13, 143-169.
- Woessmann, Ludger (2006a), Efficiency and equity of European education and training policies, CESifo Working Paper No. 1779, Munich.
- Woessmann, Ludger (2006b), Public-private partnerships and schooling outcomes across countries, CESifo Working Paper No. 1662, Munich.
- Woessmann, Ludger and Martin West (2006), Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS, *European Economic Review*, 50, 695-736.
- Wooldridge, Jeffrey (2002), *Econometric analysis of cross section and panel data*, MIT Press, Cambridge.
- Wooldridge, Jeffrey (2003), Cluster-sample methods in applied econometrics, *American Economic Review* 93(2), 133-138.

Appendix (chapter 3)

Table A3.1
Weighted means and (standard deviations) for reading

	With imp. values		Without imp. val.		Min	Max	description
	GER	FIN	GER	FIN			
Reading score	490.85 (102.77)	545.15 (87.27)	507.80 (95.70)	546.05 (87.14)	206.93	887.31	Warm estimate of reading test score
Grade level							
8 th grade	.15 (.36)	.11 (.31)	.13 (.33)	.11 (.31)	0 for other grade	1 for 8th grade	Grade level of students
9 th grade (Ref.)	.61 (.49)	.89 (.31)	.63 (.48)	.89 (.31)	0 for other grade	1 for 9th grade	Grade level of students
10 th grade	.24 (.42)	0 (0)	.24 (.43)	0 (0)	0 for other grade	1 for 10th grade	Grade level of students
Student's age	188.44 (3.37)	187.56 (3.42)	188.43 (3.40)	187.55 (3.42)	182	194	Student's age in months (-181 in regression)
Student backgr.							
Student's sex	.50 (.50)	.49 (.50)	.50 (.50)	.48 (.50)	0 for female	1 for male	Sex of students
Parents' origin	.20 (.40)	.03 (.18)	.14 (.35)	.03 (.16)	0	1 if parent foreign	Parents' place of birth
Parents no sec. ed.	.01 (.12)	.09 (.29)	.02 (.13)	.11 (.31)	0	1 for less than sec.	Highest educational level reached by a parent
Secondary ed. 2	.09 (.29)	.10 (.30)	.07 (.25)	.11 (.32)	0	1 for fin. lower second.	
Secondary ed. 3	.52 (.50)	.41 (.49)	.47 (.50)	.41 (.49)	0	1 for fin. upper second.	
Tertiary ed. (Ref.)	.38 (.48)	.40 (.49)	.45 (.50)	.37 (.48)	0	1 for fin. univers.	
Books cat. 1	.01 (.11)	.01 (.08)	.01 (.10)	.01 (.08)	0	1	No books at students home
Books cat. 2	.06 (.25)	.07 (.25)	.05 (.21)	.07 (.25)	0	1	1-10
Books cat. 3	.20 (.40)	.23 (.42)	.18 (.39)	.24 (.43)	0	1	11-50
Books cat. 4	.23 (.42)	.24 (.43)	.23 (.42)	.24 (.43)	0	1	51-100
Books cat. 5	.21 (.41)	.25 (.43)	.22 (.41)	.24 (.43)	0	1	101-250

Table A3.1 continued
Weighted means and (standard deviations) for reading

	With imp. values		Without imp. val.		Min	Max	description
	GER	FIN	GER	FIN			
Books cat. 6	.15 (.36)	.14 (.35)	.17 (.38)	.14 (.35)	0	1	251-500
Books cat. 7 (Ref.)	.12 (.33)	.06 (.25)	.15 (.35)	.06 (.24)	0	1	More than 500 books
Resources							
Teacher-student ratio	.06 (.02)	.09 (.02)	.06 (.02)	.09 (.01)	.02	.19	Teachers per students at school level
Instruction time	54.55 (4.18)	51.30 (0)	54.63 (4.40)	51.30 (0)	42.12	87.75	Minutes per year/1000
% of high educated teachers	.78 (.29)	.88 (.19)	.81 (.29)	.88 (.20)	0	1	% of teachers with highest degree
No lack of material	.92 (.28)	.92 (.27)	.93 (.26)	.90 (.29)	0	1	School lacks no material
Institutions							
Private school	.04 (.20)	.03 (.17)	.04 (.19)	.03 (.18)	0 if school is public	1	School type
Standardized tests	.03 (.18)	.26 (.44)	.03 (.17)	.26 (.44)	0	1	Standardized tests more than once a year
Selection of students	.66 (.47)	.20 (.40)	.59 (.49)	.19 (.40)	0 if school may not select	1	School has right to select its students
Budget (category variable)	1.07 (.39)	1.55 (.52)	1.08 (.40)	1.56 (.53)	0	2	School's right over budget allocation and formulation
School types							
Vocational school	.05 (.22)	0	.03 (.18)	0	0	1	Vocational school (Berufsschule)
Low sec. school	.19 (.39)	0	.19 (.39)	0	0	1	Low sec. school (Hauptschule)
Medium sec. school	.26 (.44)	0	.29 (.45)	0	0	1	Medium sec. school (Realschule)
Highest sec. school	.29 (.45)	0	.35 (.48)	0	0	1	Highest sec. school (Gymnasium)
Comprehensive school (Ref.)	.17 (.38)	1	.14 (.35)	1	0	1	Comprehensive school (Gesamtschule)
School type n.a.	.03 (.17)	0	0	0	0	1	No information on school type

Table A3.2
Weighted means of selected variables by school type for Germany

School type	Students (in %)	Reading score	Parents tert. educ.	% of high educ. teachers	Private school	Teacher- student ratio
all	4,917 (100)	490.85 (102.77)	.38 (.48)	.78 (.29)	.04 (.20)	.06 (.02)
Vocational	116 (2.4)	476.15 (72.71)	.28 (.45)	.77 (.28)	0 (0)	.05 (.01)
Low second.	931 (18.9)	405.96 (79.86)	.18 (.39)	.54 (.31)	0 (0)	.06 (.01)
Medium second.	1,235 (25.1)	498.19 (73.10)	.34 (.47)	.81 (.27)	.05 (.22)	.06 (.03)
High second.	1,713 (34.8)	577.51 (68.84)	.61 (.49)	.97 (.10)	.09 (.29)	.06 (.01)
Comprehensive	885 (18.0)	467.02 (84.81)	.32 (.47)	.79 (.27)	0 (0)	.06 (.01)
School type n.a.	37 (0.8)	287.77 (70.65)	.21 (.41)	.28 (.17)	0 (0)	.10 (.03)

Standard deviations are reported in parentheses. Only in first column the percentage of students are reported in parentheses.

Table A3.3
Age and reading test scores by grade level

	GER		FIN	
	Obs.	Mean (SD)	Obs.	Mean (SD)
Age				
Grade 8	682	187.08 (2.86)	530	183.29 (2.63)
Grade 9	3,131	187.44 (3.03)	4,325	188.10 (3.13)
Grade 10	1,104	191.93 (1.77)	0	-
Reading test score				
Grade 8	682	409.71 (99.95)	530	497.96 (93.74)
Grade 9	3,131	492.23 (95.49)	4,325	550.96 (84.64)
Grade 10	1,104	539.87 (89.52)	0	-

Number of observations and weighted means. Standard deviations in parentheses.

Table A3.4
Coefficients (standard errors) of weighted survey regressions

	With imputed values		Without imp. values		Percentage of missing values	
	GER	FIN	GER	FIN	GER	FIN
Grade level						
8th grade	-37.75*** (3.83)	-47.68*** (4.93)	-43.26*** (4.44)	-42.80*** (5.79)	0	0
10th grade	48.70*** (3.42)	-	48.76*** (4.20)	-	0	-
Student's age	-2.65*** (.41)	-.76** (.33)	-2.58*** (.50)	-.46 (.41)	0	0
Student background						
Student's sex	-10.34*** (2.54)	-44.94*** (2.50)	-14.68*** (3.06)	-45.88*** (2.89)	0	0
Parents' origin	-13.90*** (3.62)	-32.07*** (7.30)	-13.27*** (4.51)	-40.80*** (9.99)	2.01	1.40
Parents no sec. ed.	-34.78*** (9.45)	-26.36*** (4.13)	-33.75*** (11.61)	-30.22*** (4.70)	16.11	10.24
Secondary ed. 2	-21.39*** (4.92)	-30.45*** (4.36)	-23.90*** (6.21)	-34.05*** (4.74)		
Secondary ed. 3	-1.85 (2.80)	-10.90*** (3.07)	-2.45 (2.87)	-12.53*** (3.37)		
Books cat. 1	-64.71*** (10.98)	-68.11*** (17.70)	-63.96*** (15.33)	-59.26*** (14.99)		
Books cat. 2	-57.70*** (8.10)	-49.67*** (7.60)	-50.55*** (6.96)	-50.64*** (10.12)		
Books cat. 3	-28.80*** (4.49)	-35.60*** (5.86)	-27.74*** (5.31)	-32.33*** (7.40)		
Books cat. 4	-18.94*** (4.24)	-29.39*** (5.62)	-18.87*** (5.23)	-27.33*** (7.41)	1.87	1.36
Books cat. 5	-11.05*** (4.00)	-9.32 (5.75)	-13.60*** (4.82)	-5.17 (7.21)		
Books cat. 6	-8.39** (4.10)	-.30 (5.80)	-8.69* (4.79)	.70 (7.71)		
Resources						
Teacher-student ratio	-15.67 (86.54)	-96.69 (122.58)	47.14 (107.03)	-51.97 (131.81)	14.46	10.61
Instruction time	-.52 (0.47)	-	.10 (0.50)	-	14.79	0
% of high educated teachers	7.77 (8.70)	-8.61 (9.24)	10.50 (10.98)	2.80 (9.83)	8.24	14.56
No lack of material	6.62 (6.18)	6.51 (8.77)	-.95 (7.59)	-2.14 (6.73)	8.56	1.11

Table A3.4 continued
Coefficients (standard errors) of weighted survey regressions

	With imputed values		Without imp. values		Percentage of missing values	
	GER	FIN	GER	FIN	GER	FIN
Institutions						
Private school	4.91 (9.00)	2.48 (10.73)	-3.39 (12.89)	-4.89 (12.76)	8.77	0
Standardized tests	-15.92 (12.44)	2.12 (3.83)	-15.68 (12.10)	4.38 (3.72)	10.86	0
Selection of students	11.10*** (4.21)	-7.05 (4.35)	18.08*** (5.00)	2.39 (4.95)	8.79	3.36
Budget (category variable)	-1.04 (4.33)	-6.25 (3.88)	-1.57 (5.58)	-10.83*** (3.88)	8.24	0.66
School types						
Voc. school	-24.13*** (8.08)	-	-45.83*** (14.91)	-	-	-
Low sec. school	-33.51** (6.85)	-	-33.24*** (8.77)	-	-	-
Medium sec. school	27.99*** (5.91)	-	18.60*** (7.52)	-	-	-
Higher sec. school	85.31*** (5.83)	-	74.39*** (7.23)	-	-	-
School type n.a.	-146.91*** (10.39)	-	-	-	-	-
Intercept	525.79*** (27.53)	627.78*** (16.11)	502.18*** (31.47)	628.83*** (15.53)		
Number of observat.	4917	4855	3080	3407		
R-squared	.5566	.1747	.5074	.1800		
F-Test	224.85	42.01	67.97	32.26		

P-Values: *** 1 Percent
 ** 5 Percent
 * 10 Percent

Cluster-robust standard errors are reported in parentheses.

Table A3.5
Decomposition for significantly different coefficients at ten-percent-level

	Sum	Grade	Student Background	Resources	Institutions	Schools	Cons
Total gap	54.29						
Charact. Effect	-40.60	-6.38	-1.14	0.56	-7.67	-25.98	
Ratios	-0.75	-0.12	-0.02	0.01	-0.14	-0.48	
Return Effect	56.07	6.85	-25.80	0	0	-25.98	101.00
Ratios	1.03	0.13	-0.48	0.00	0.00	-0.48	1.86
Interaction effect	38.83	8.31	4.54	0	0	25.98	
Ratios	0.72	0.15	0.08	0.00	0.00	0.48	

Oaxaca-Blinder decomposition of test score gap between Finland and Germany. Germany is country of reference. Decomposition includes additional interaction effect (see equation (3.3)). Explanatory variables are divided into five groups. Only those variables are used for the decomposition whose coefficients differ between the countries at the ten-percent-significance level. Absolute effects in reading scores. Ratios are effects divided by total score gap.

Figure A3.1
Real and hypothetical test score distributions

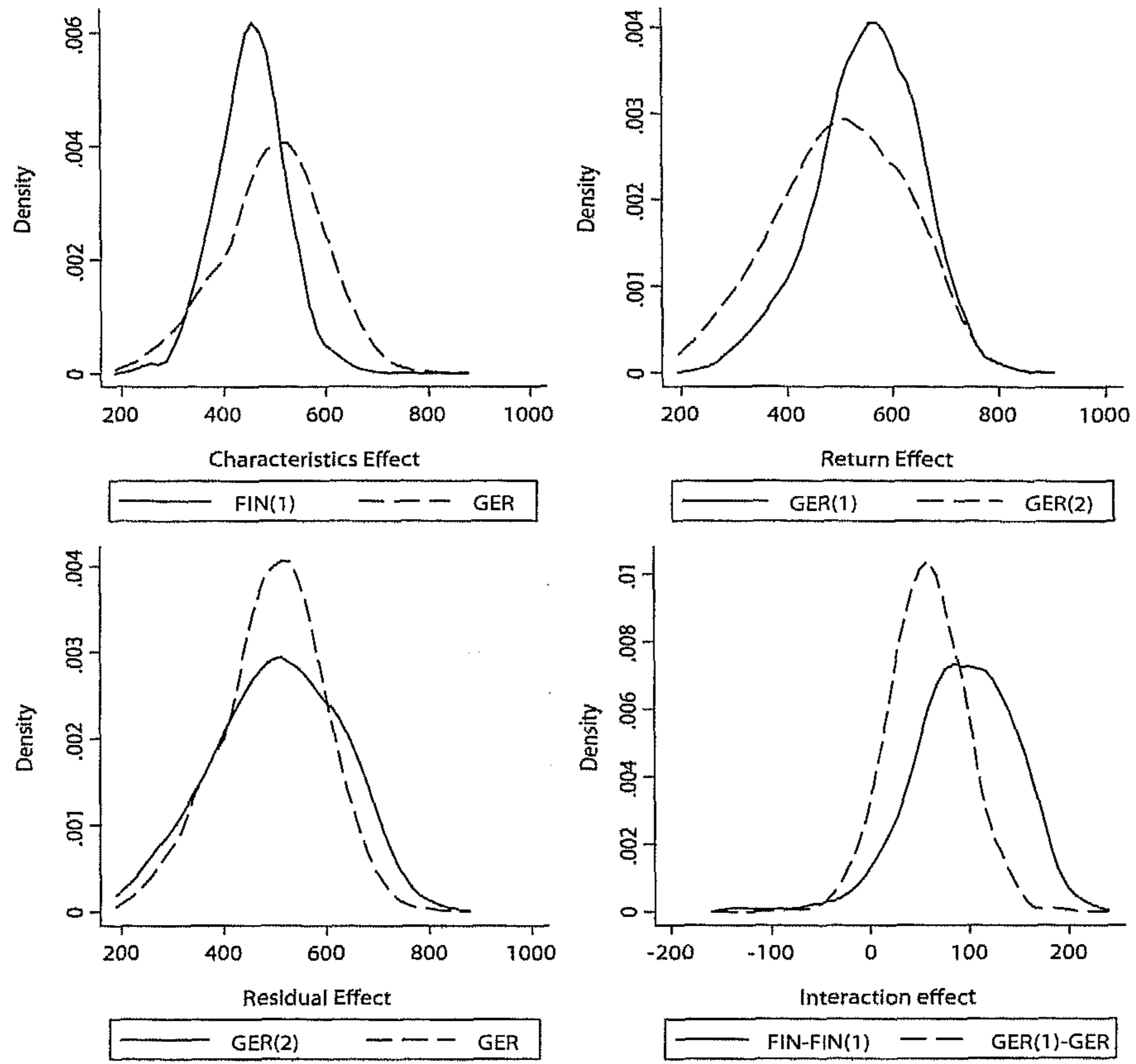


Figure A3.2
Total effects with confidence intervals

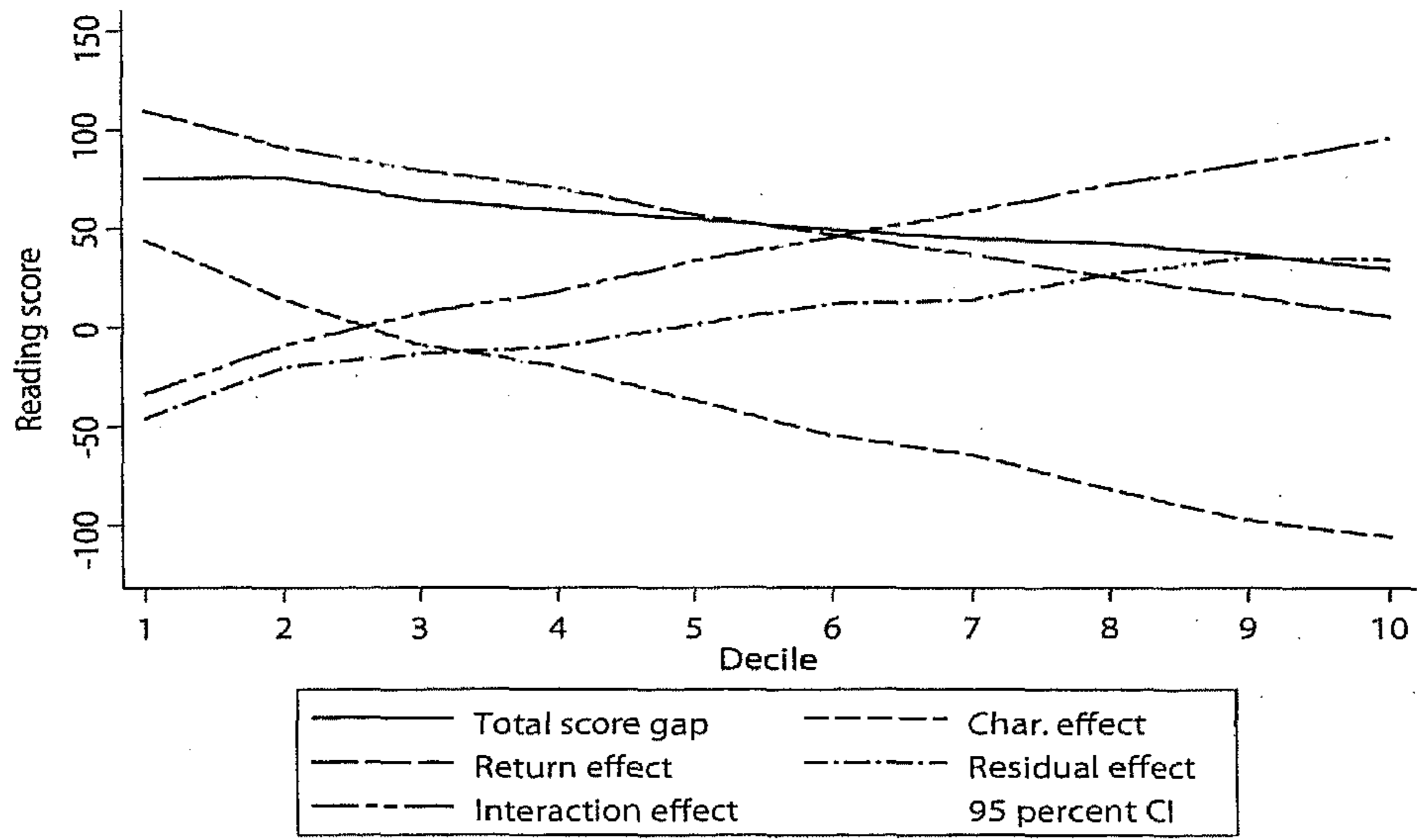


Figure A3.3
Components of the characteristics effect with confidence intervals

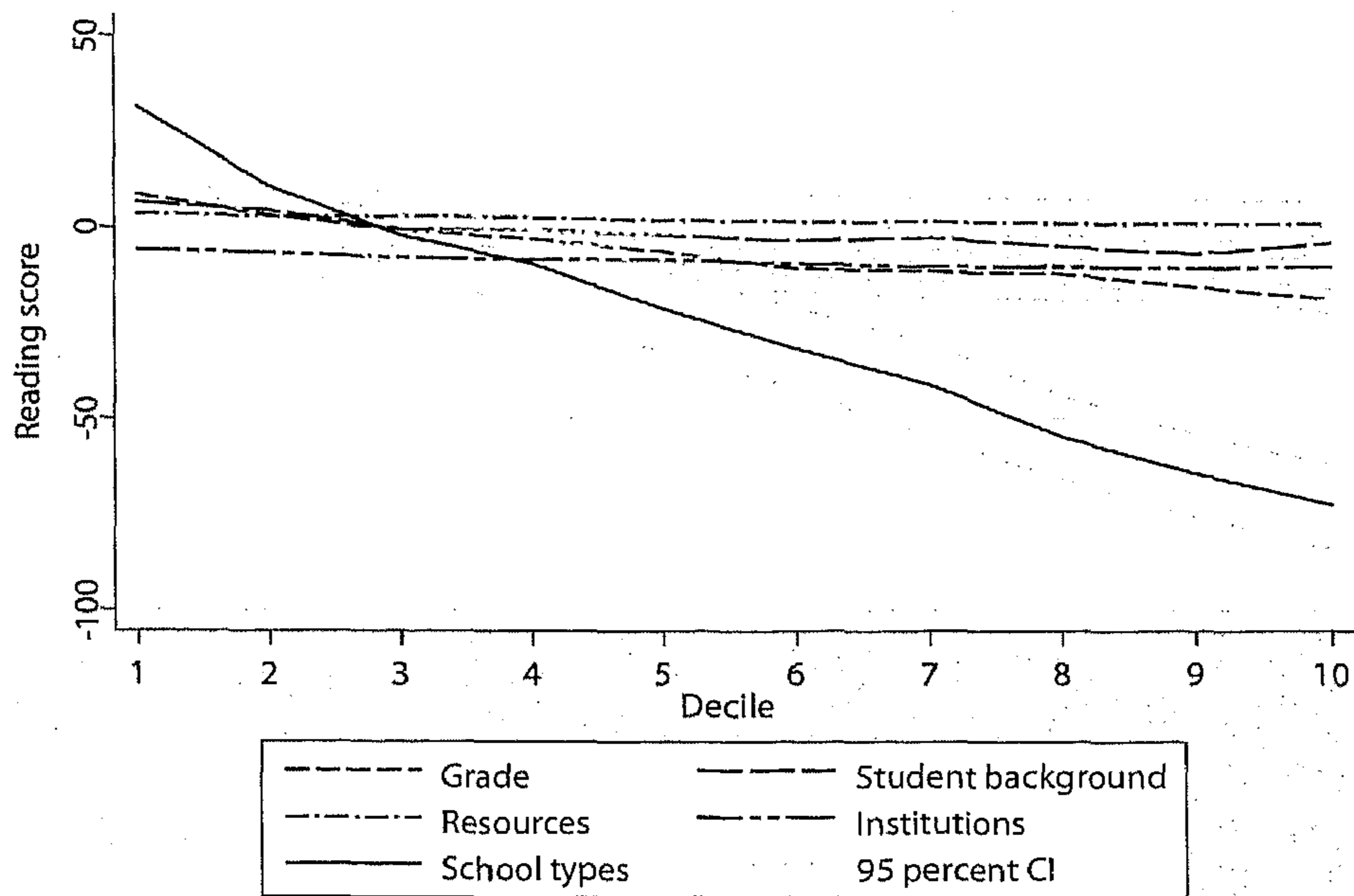
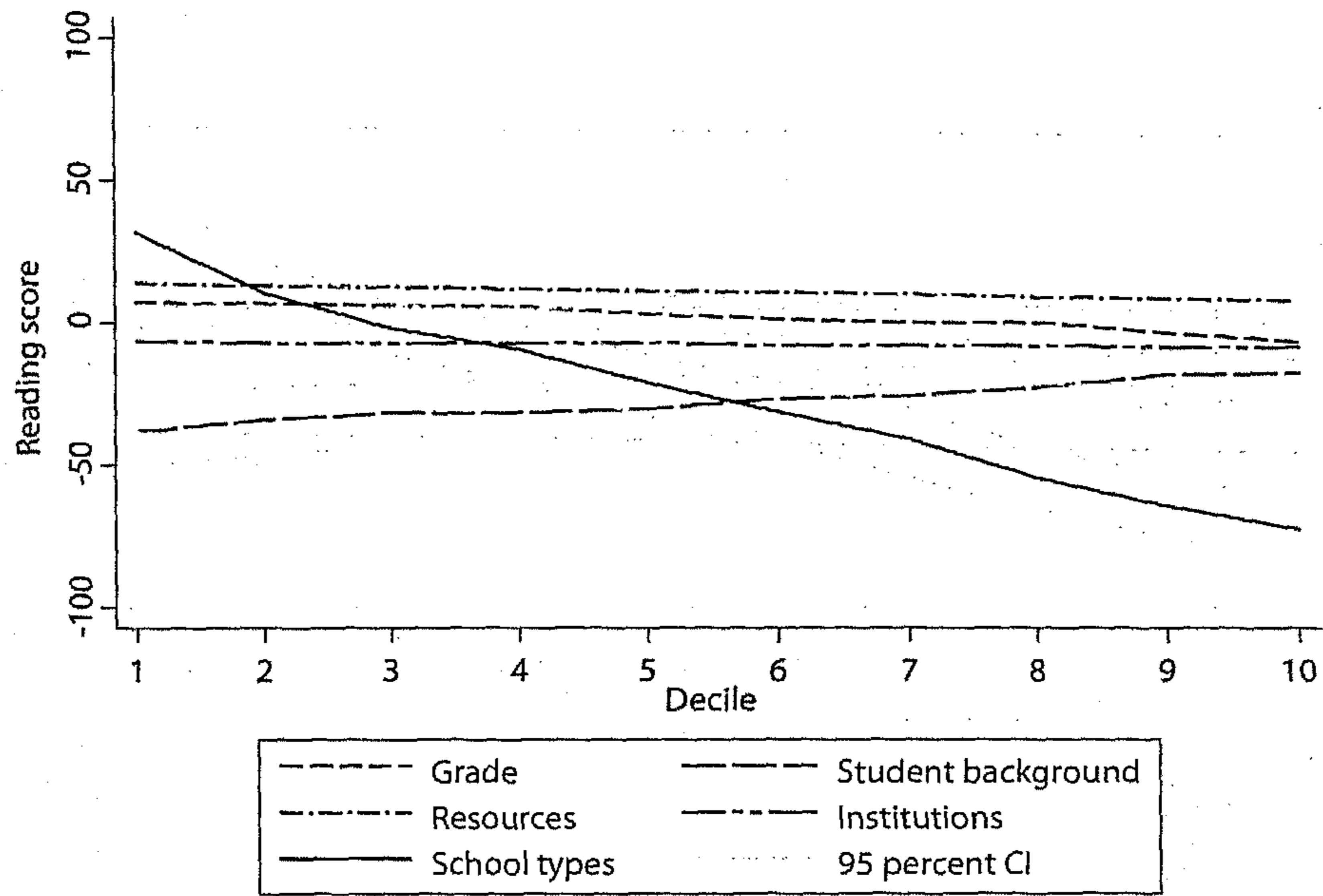


Figure A3.4
 Components of the return effects with confidence intervals



Appendix (chapter 4)

Table A4.1
Participation of students and schools in PIRLS and PISA

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
PIRLS														
Number of students	8,177	3,005	7,515	3,126	3,503	2,478	4,650	3,635	3,502	3,014	3,388	2,458	4,092	6,027
Number of schools	204	141	211	131	145	145	216	133	184	140	136	156	206	146
PISA														
Number of students	29,005	5,170	4,981	9,228	4,642	4,580	4,808	3,193	4,782	3,842	4,062	3,632	6,686	4,362
Number of schools	1,118	221	220	363	178	158	195	131	173	155	177	154	247	155

Table A4.2
Means and standard deviations of the data including imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Scores														
PIRLS	544.7 (67.7)	537.2 (60.6)	539.7 (63.2)	553.3 (82.8)	525.9 (66.1)	524.7 (69.2)	543.6 (61.8)	513.2 (70.7)	540.7 (67.0)	544.7 (57.9)	500.3 (77.1)	530.2 (89.2)	527.9 (62.7)	561.1 (61.6)
PISA	536.4 (90.5)	501.0 (83.1)	485.1 (108.3)	524.5 (97.2)	505.2 (88.7)	474.7 (93.8)	481.3 (90.3)	510.7 (87.1)	490.5 (86.9)	459.1 (98.7)	507.4 (99.2)	530.4 (104.1)	462.0 (88.8)	517.0 (88.6)
Student Background Variables														
PIRLS														
Female	0.50 (0.50)	0.49 (0.50)	0.50 (0.50)	0.52 (0.50)	0.49 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.48 (0.50)	0.48 (0.50)	0.48 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)
Age	10.02 (0.41)	10.51 (0.45)	10.54 (0.50)	10.21 (0.30)	10.12 (0.50)	9.95 (0.44)	10.67 (0.52)	9.72 (0.29)	9.85 (0.37)	11.05 (0.48)	9.97 (0.33)	10.05 (0.36)	10.28 (0.63)	10.80 (0.32)
Parents' origin	0.35 (0.48)	0.11 (0.32)	0.23 (0.42)	0.32 (0.47)	0.28 (0.45)	0.19 (0.39)	0.06 (0.24)	0.12 (0.33)	0.11 (0.32)	0.33 (0.47)	0.17 (0.38)	0.38 (0.49)	0.23 (0.42)	0.24 (0.43)
Language	0.16 (0.37)	0.07 (0.25)	0.10 (0.30)	0.12 (0.32)	0.13 (0.34)	0.06 (0.23)	0.04 (0.21)	0.11 (0.32)	0.04 (0.19)	0.08 (0.28)	0.08 (0.27)	0.16 (0.37)	0.15 (0.35)	0.10 (0.30)
Books at home														
< 11	0.05 (0.22)	0.01 (0.11)	0.06 (0.23)	0.04 (0.19)	0.10 (0.29)	0.09 (0.29)	0.04 (0.20)	0.01 (0.11)	0.11 (0.32)	0.03 (0.17)	0.02 (0.13)	0.05 (0.22)	0.10 (0.30)	0.03 (0.17)
11-25	0.10 (0.30)	0.07 (0.25)	0.11 (0.31)	0.06 (0.24)	0.13 (0.34)	0.14 (0.35)	0.07 (0.25)	0.04 (0.21)	0.20 (0.40)	0.07 (0.25)	0.05 (0.22)	0.10 (0.29)	0.16 (0.37)	0.06 (0.24)
26-100	0.48 (0.50)	0.42 (0.49)	0.42 (0.49)	0.63 (0.48)	0.40 (0.49)	0.51 (0.50)	0.27 (0.44)	0.23 (0.42)	0.38 (0.49)	0.32 (0.47)	0.25 (0.43)	0.45 (0.50)	0.36 (0.48)	0.24 (0.43)
101-200	0.17 (0.37)	0.21 (0.41)	0.16 (0.36)	0.12 (0.32)	0.16 (0.36)	0.12 (0.32)	0.21 (0.41)	0.22 (0.41)	0.13 (0.34)	0.21 (0.41)	0.20 (0.40)	0.17 (0.37)	0.17 (0.38)	0.20 (0.40)
>200	0.20 (0.40)	0.29 (0.45)	0.26 (0.44)	0.16 (0.36)	0.22 (0.41)	0.14 (0.35)	0.41 (0.49)	0.50 (0.50)	0.16 (0.37)	0.37 (0.48)	0.48 (0.50)	0.24 (0.42)	0.21 (0.41)	0.46 (0.50)
Parents' education														
No Second.	0.03 (0.16)	0.03 (0.18)	0.09 (0.28)	0.65 (0.48)	0.42 (0.49)	0.14 (0.35)	0.08 (0.28)	0.11 (0.32)	0.29 (0.46)	0.04 (0.28)	0.03 (0.16)	0.03 (0.18)	0.04 (0.19)	0.06 (0.23)
Upper Second.	0.24 (0.42)	0.69 (0.46)	0.20 (0.40)	0.09 (0.28)	0.16 (0.37)	0.39 (0.49)	0.57 (0.49)	0.36 (0.48)	0.48 (0.50)	0.37 (0.48)	0.42 (0.49)	0.39 (0.49)	0.26 (0.44)	0.34 (0.48)
Post-Second.	0.46 (0.50)	0.06 (0.24)	0.46 (0.50)	0.02 (0.13)	0.13 (0.33)	0.19 (0.39)	0.03 (0.17)	0.08 (0.28)	0.04 (0.18)	0.31 (0.46)	---	0.18 (0.39)	0.43 (0.49)	0.21 (0.41)
University	0.26 (0.44)	0.21 (0.41)	0.24 (0.43)	0.20 (0.40)	0.21 (0.41)	0.20 (0.40)	0.31 (0.46)	0.44 (0.50)	0.17 (0.38)	0.28 (0.45)	0.55 (0.50)	0.38 (0.48)	0.27 (0.44)	0.39 (0.49)
School location														
City	0.35 (0.48)	0.15 (0.36)	0.23 (0.42)	0.30 (0.46)	0.15 (0.35)	0.29 (0.45)	0.33 (0.47)	0.27 (0.45)	0.16 (0.37)	0.42 (0.49)	0.19 (0.39)	0.45 (0.50)	0.39 (0.49)	0.16 (0.37)
Town	0.56 (0.50)	0.73 (0.44)	0.75 (0.43)	0.68 (0.47)	0.78 (0.41)	0.66 (0.47)	0.62 (0.48)	0.60 (0.49)	0.80 (0.40)	0.51 (0.50)	0.66 (0.47)	0.46 (0.50)	0.61 (0.49)	0.77 (0.42)
Rural	0.09 (0.28)	0.11 (0.31)	0.01 (0.12)	0.02 (0.14)	0.07 (0.26)	0.05 (0.21)	0.05 (0.21)	0.13 (0.34)	0.04 (0.19)	0.07 (0.25)	0.15 (0.36)	0.09 (0.28)	---	0.07 (0.25)
Parents' attitude														
High	0.70 (0.46)	0.69 (0.46)	0.52 (0.50)	0.82 (0.38)	0.48 (0.50)	0.68 (0.47)	0.75 (0.43)	0.72 (0.45)	0.57 (0.50)	0.48 (0.50)	0.76 (0.43)	0.70 (0.46)	0.51 (0.50)	0.74 (0.42)
Medium	0.27 (0.45)	0.28 (0.45)	0.39 (0.49)	0.14 (0.35)	0.48 (0.50)	0.27 (0.44)	0.23 (0.43)	0.25 (0.43)	0.38 (0.48)	0.50 (0.50)	0.21 (0.41)	0.26 (0.44)	0.45 (0.50)	0.22 (0.41)
Low	0.03 (0.17)	0.03 (0.18)	0.09 (0.28)	0.03 (0.17)	0.04 (0.19)	0.05 (0.22)	0.02 (0.15)	0.03 (0.16)	0.06 (0.23)	0.02 (0.15)	0.04 (0.19)	0.04 (0.20)	0.05 (0.21)	0.05 (0.21)

Table A4.2 continued

Means and standard deviations of the data including imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
PISA														
Female	0.50 (0.50)	0.53 (0.50)	0.50 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.51 (0.50)	0.49 (0.50)	0.50 (0.50)	0.50 (0.50)	0.49 (0.50)
Age	15.79 (0.28)	15.72 (0.28)	15.70 (0.28)	15.62 (0.29)	15.79 (0.28)	15.70 (0.27)	15.71 (0.28)	15.63 (0.28)	15.72 (0.28)	15.72 (0.28)	15.73 (0.28)	15.70 (0.28)	15.70 (0.28)	15.72 (0.28)
Parents' origin	0.31 (0.46)	0.08 (0.27)	0.21 (0.40)	0.18 (0.38)	0.25 (0.43)	0.10 (0.31)	0.04 (0.19)	0.06 (0.24)	0.05 (0.22)	0.39 (0.49)	0.11 (0.31)	0.36 (0.48)	0.13 (0.34)	0.21 (0.41)
Language	0.13 (0.34)	0.01 (0.12)	0.16 (0.37)	0.05 (0.22)	0.08 (0.27)	0.04 (0.19)	---	0.02 (0.15)	0.22 (0.41)	0.09 (0.28)	0.13 (0.33)	0.14 (0.35)	0.08 (0.27)	0.10 (0.30)
Books at home														
< 11	0.06 (0.24)	0.02 (0.16)	0.08 (0.27)	0.08 (0.28)	0.11 (0.31)	0.11 (0.31)	0.04 (0.21)	0.03 (0.17)	0.09 (0.29)	0.04 (0.19)	0.06 (0.24)	0.06 (0.24)	0.07 (0.25)	0.06 (0.23)
11-25	0.18 (0.38)	0.11 (0.32)	0.20 (0.40)	0.21 (0.41)	0.23 (0.42)	0.30 (0.46)	0.13 (0.34)	0.11 (0.31)	0.21 (0.41)	0.10 (0.30)	0.13 (0.34)	0.18 (0.38)	0.17 (0.37)	0.14 (0.35)
26-100	0.20 (0.40)	0.21 (0.41)	0.22 (0.41)	0.21 (0.41)	0.21 (0.41)	0.26 (0.44)	0.19 (0.39)	0.20 (0.40)	0.23 (0.42)	0.19 (0.39)	0.18 (0.39)	0.21 (0.41)	0.20 (0.40)	0.18 (0.39)
101-200	0.24 (0.43)	0.28 (0.45)	0.21 (0.41)	0.21 (0.40)	0.20 (0.40)	0.16 (0.37)	0.22 (0.41)	0.25 (0.43)	0.21 (0.41)	0.27 (0.44)	0.23 (0.42)	0.23 (0.42)	0.22 (0.41)	0.24 (0.43)
>200	0.32 (0.47)	0.37 (0.48)	0.29 (0.45)	0.29 (0.45)	0.24 (0.43)	0.16 (0.37)	0.42 (0.49)	0.41 (0.49)	0.27 (0.44)	0.41 (0.49)	0.39 (0.49)	0.32 (0.47)	0.34 (0.47)	0.37 (0.48)
Parents' education														
No Second.	0.06 (0.24)	0.01 (0.10)	0.05 (0.22)	0.06 (0.24)	0.15 (0.35)	0.15 (0.35)	0.06 (0.23)	0.12 (0.32)	0.27 (0.44)	0.03 (0.17)	0.06 (0.24)	0.05 (0.21)	0.02 (0.15)	0.06 (0.24)
Upper Second.	---	0.30 (0.46)	0.46 (0.50)	0.33 (0.47)	0.20 (0.40)	0.05 (0.21)	0.31 (0.46)	0.25 (0.44)	0.12 (0.33)	0.03 (0.17)	0.13 (0.33)	0.25 (0.43)	0.02 (0.14)	0.08 (0.26)
Post-Second.	0.26 (0.44)	0.45 (0.50)	0.09 (0.29)	0.08 (0.26)	0.13 (0.34)	0.23 (0.42)	0.35 (0.48)	0.20 (0.40)	0.36 (0.48)	0.49 (0.50)	0.19 (0.40)	0.06 (0.25)	0.51 (0.50)	0.22 (0.41)
University	0.67 (0.47)	0.24 (0.43)	0.38 (0.49)	0.52 (0.50)	0.49 (0.50)	0.45 (0.50)	0.28 (0.45)	0.37 (0.48)	0.21 (0.41)	0.44 (0.50)	0.60 (0.49)	0.63 (0.48)	0.44 (0.50)	0.63 (0.48)
School location														
City	---	0.30 (0.46)	0.21 (0.41)	0.33 (0.47)	0.17 (0.38)	0.32 (0.47)	0.42 (0.49)	---	0.29 (0.46)	0.30 (0.46)	0.11 (0.31)	0.46 (0.50)	0.37 (0.48)	0.17 (0.38)
Town	---	0.61 (0.49)	0.72 (0.45)	0.60 (0.49)	0.77 (0.42)	0.63 (0.48)	0.56 (0.50)	---	0.69 (0.46)	0.49 (0.50)	0.51 (0.50)	0.42 (0.49)	0.32 (0.47)	0.61 (0.49)
Rural	---	0.09 (0.28)	0.06 (0.25)	0.08 (0.27)	0.06 (0.23)	0.05 (0.21)	0.02 (0.15)	---	0.02 (0.14)	0.20 (0.40)	0.39 (0.49)	0.13 (0.33)	0.31 (0.46)	0.22 (0.41)
Parents' attitude														
High	0.15 (0.36)	0.09 (0.29)	0.09 (0.28)	0.13 (0.33)	0.20 (0.40)	0.15 (0.36)	0.23 (0.42)	0.17 (0.38)	0.27 (0.44)	0.22 (0.41)	0.10 (0.29)	0.15 (0.36)	0.22 (0.42)	0.09 (0.29)
Medium	0.70 (0.46)	0.69 (0.46)	0.71 (0.45)	0.74 (0.44)	0.70 (0.46)	0.74 (0.44)	0.67 (0.47)	0.67 (0.47)	0.63 (0.48)	0.66 (0.47)	0.68 (0.46)	0.70 (0.46)	0.63 (0.48)	0.71 (0.45)
Low	0.15 (0.35)	0.21 (0.41)	0.20 (0.40)	0.14 (0.34)	0.10 (0.30)	0.11 (0.31)	0.09 (0.29)	0.16 (0.36)	0.10 (0.30)	0.12 (0.32)	0.22 (0.41)	0.15 (0.36)	0.15 (0.36)	0.19 (0.40)

Values are weighted by the sampling probability of students. Standard deviations are reported in parentheses.

Table A4.3
Means and standard deviations of the data without imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Scores														
PIRLS	555.8 (65.3)	546.1 (57.2)	547.9 (61.8)	576.0 (82.0)	534.9 (63.4)	537.8 (63.6)	557.4 (57.3)	519.6 (69.4)	546.8 (65.6)	562.1 (51.6)	507.7 (76.0)	538.5 (87.4)	543.0 (58.3)	564.5 (61.2)
PISA	501.2 (106)	504.6 (81.1)	512.4 (90.7)	537.7 (92.4)	516.2 (83.2)	479.2 (91.3)	482.0 (86.6)	511.2 (86.8)	497.2 (83.9)	469.2 (95.9)	517.2 (94.5)	550.6 (96.0)	468.9 (86.7)	523.0 (86.0)
Student background variables														
PIRLS														
Female	0.50 (0.50)	0.49 (0.50)	0.50 (0.50)	0.52 (0.50)	0.49 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.48 (0.50)	0.48 (0.50)	0.48 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)
Age	10.02 (0.41)	10.51 (0.45)	10.54 (0.50)	10.21 (0.30)	10.12 (0.50)	9.95 (0.44)	10.67 (0.52)	9.72 (0.29)	9.85 (0.37)	11.05 (0.48)	9.97 (0.33)	10.05 (0.36)	10.28 (0.63)	10.80 (0.32)
Parents' origin	0.39 (0.49)	0.12 (0.32)	0.25 (0.43)	0.33 (0.47)	0.30 (0.46)	0.19 (0.39)	0.07 (0.25)	0.13 (0.34)	0.12 (0.32)	0.35 (0.48)	0.18 (0.38)	0.41 (0.49)	0.24 (0.43)	0.24 (0.43)
Language	0.16 (0.37)	0.07 (0.25)	0.10 (0.31)	0.12 (0.32)	0.13 (0.34)	0.06 (0.24)	0.04 (0.21)	0.12 (0.32)	0.04 (0.19)	0.09 (0.28)	0.08 (0.28)	0.16 (0.37)	0.15 (0.35)	0.10 (0.30)
Books at home														
< 11	0.06 (0.24)	0.01 (0.12)	0.06 (0.24)	0.07 (0.25)	0.10 (0.30)	0.10 (0.30)	0.04 (0.20)	0.01 (0.11)	0.12 (0.32)	0.03 (0.18)	0.02 (0.14)	0.06 (0.24)	0.10 (0.30)	0.03 (0.18)
11-25	0.12 (0.33)	0.08 (0.26)	0.12 (0.33)	0.11 (0.31)	0.14 (0.35)	0.16 (0.37)	0.07 (0.26)	0.05 (0.22)	0.21 (0.41)	0.07 (0.26)	0.05 (0.23)	0.11 (0.32)	0.16 (0.37)	0.07 (0.25)
26-100	0.37 (0.48)	0.37 (0.48)	0.36 (0.48)	0.34 (0.47)	0.35 (0.48)	0.44 (0.50)	0.27 (0.44)	0.27 (0.44)	0.36 (0.48)	0.32 (0.47)	0.26 (0.44)	0.35 (0.48)	0.35 (0.48)	0.25 (0.43)
101-200	0.20 (0.40)	0.24 (0.42)	0.18 (0.38)	0.21 (0.41)	0.17 (0.38)	0.13 (0.34)	0.22 (0.40)	0.26 (0.44)	0.14 (0.35)	0.23 (0.42)	0.22 (0.42)	0.21 (0.40)	0.17 (0.38)	0.22 (0.42)
>200	0.24 (0.43)	0.30 (0.46)	0.27 (0.45)	0.28 (0.45)	0.23 (0.42)	0.16 (0.37)	0.40 (0.49)	0.40 (0.49)	0.17 (0.38)	0.35 (0.48)	0.45 (0.50)	0.28 (0.45)	0.21 (0.41)	0.42 (0.50)
Parents' education														
No Second.	0.03 (0.18)	0.04 (0.20)	0.12 (0.32)	0.35 (0.48)	0.31 (0.46)	0.17 (0.37)	0.08 (0.28)	0.13 (0.34)	0.30 (0.46)	0.05 (0.21)	0.03 (0.17)	0.04 (0.20)	0.04 (0.19)	0.06 (0.24)
Upper Second.	0.26 (0.44)	0.66 (0.47)	0.23 (0.42)	0.17 (0.38)	0.20 (0.40)	0.28 (0.45)	0.57 (0.50)	0.42 (0.49)	0.46 (0.50)	0.35 (0.48)	0.44 (0.49)	0.34 (0.47)	0.27 (0.44)	0.36 (0.48)
Post-Second.	0.37 (0.48)	0.07 (0.26)	0.36 (0.48)	0.03 (0.18)	0.16 (0.37)	0.23 (0.42)	0.03 (0.18)	0.10 (0.30)	0.04 (0.19)	0.34 (0.47)	---	0.22 (0.42)	0.42 (0.49)	0.23 (0.42)
University	0.32 (0.47)	0.22 (0.42)	0.27 (0.44)	0.36 (0.48)	0.23 (0.42)	0.23 (0.42)	0.30 (0.46)	0.33 (0.47)	0.18 (0.38)	0.26 (0.44)	0.53 (0.50)	0.37 (0.48)	0.27 (0.44)	0.35 (0.48)
School location														
City	0.37 (0.48)	0.19 (0.39)	0.31 (0.46)	0.38 (0.49)	0.21 (0.41)	0.38 (0.48)	0.42 (0.49)	0.35 (0.48)	0.16 (0.37)	0.44 (0.50)	0.24 (0.43)	0.45 (0.50)	0.56 (0.50)	0.19 (0.40)
Town	0.53 (0.50)	0.69 (0.47)	0.66 (0.47)	0.59 (0.49)	0.69 (0.46)	0.56 (0.50)	0.52 (0.50)	0.47 (0.50)	0.80 (0.40)	0.45 (0.50)	0.56 (0.50)	0.44 (0.50)	0.44 (0.50)	0.73 (0.44)
Rural	0.10 (0.30)	0.14 (0.35)	0.02 (0.14)	0.03 (0.17)	0.10 (0.30)	0.06 (0.24)	0.06 (0.24)	0.18 (0.39)	0.04 (0.19)	0.11 (0.31)	0.20 (0.40)	0.11 (0.31)	---	0.08 (0.27)
Parents' attitude														
High	0.64 (0.48)	0.64 (0.48)	0.55 (0.50)	0.69 (0.46)	0.53 (0.50)	0.64 (0.48)	0.74 (0.44)	0.67 (0.47)	0.56 (0.50)	0.48 (0.50)	0.73 (0.44)	0.65 (0.48)	0.51 (0.50)	0.71 (0.45)
Medium	0.32 (0.47)	0.32 (0.47)	0.35 (0.48)	0.26 (0.44)	0.43 (0.50)	0.31 (0.46)	0.23 (0.43)	0.30 (0.46)	0.37 (0.48)	0.50 (0.50)	0.23 (0.42)	0.30 (0.46)	0.45 (0.50)	0.24 (0.43)
Low	0.04 (0.19)	0.04 (0.20)	0.10 (0.30)	0.06 (0.23)	0.04 (0.21)	0.06 (0.23)	0.03 (0.16)	0.03 (0.18)	0.07 (0.25)	0.03 (0.16)	0.04 (0.20)	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)

Table A4.3 continued
Means and standard deviations of the data without imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
PISA														
Female	0.50 (0.50)	0.53 (0.50)	0.50 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.51 (0.50)	0.49 (0.50)	0.50 (0.50)	0.50 (0.50)	0.49 (0.50)
Age	15.79 (0.28)	15.72 (0.28)	15.70 (0.28)	15.62 (0.29)	15.79 (0.28)	15.70 (0.27)	15.71 (0.28)	15.63 (0.28)	15.72 (0.28)	15.72 (0.28)	15.73 (0.28)	15.70 (0.28)	15.70 (0.28)	15.72 (0.28)
Parents' origin	0.31 (0.46)	0.08 (0.27)	0.21 (0.41)	0.18 (0.38)	0.25 (0.43)	0.11 (0.31)	0.04 (0.19)	0.06 (0.24)	0.05 (0.22)	0.40 (0.49)	0.11 (0.31)	0.37 (0.48)	0.13 (0.33)	0.21 (0.41)
Language	0.11 (0.32)	0.01 (0.07)	0.08 (0.27)	0.04 (0.20)	0.05 (0.22)	0.03 (0.16)	---	0.02 (0.13)	0.18 (0.38)	0.07 (0.25)	0.06 (0.24)	0.10 (0.30)	0.07 (0.26)	0.07 (0.26)
Books at home														
< 11	0.06 (0.24)	0.03 (0.16)	0.08 (0.28)	0.09 (0.28)	0.12 (0.32)	0.11 (0.32)	0.05 (0.21)	0.03 (0.17)	0.09 (0.29)	0.04 (0.19)	0.07 (0.25)	0.06 (0.24)	0.07 (0.26)	0.06 (0.23)
11-25	0.18 (0.38)	0.11 (0.32)	0.20 (0.40)	0.21 (0.41)	0.22 (0.41)	0.28 (0.45)	0.14 (0.34)	0.11 (0.31)	0.21 (0.41)	0.10 (0.30)	0.14 (0.35)	0.17 (0.38)	0.17 (0.38)	0.14 (0.35)
26-100	0.20 (0.40)	0.21 (0.41)	0.23 (0.42)	0.21 (0.41)	0.22 (0.42)	0.27 (0.44)	0.20 (0.40)	0.21 (0.41)	0.23 (0.42)	0.19 (0.39)	0.19 (0.39)	0.22 (0.41)	0.20 (0.40)	0.18 (0.39)
101-200	0.24 (0.43)	0.28 (0.45)	0.21 (0.41)	0.21 (0.41)	0.22 (0.41)	0.17 (0.38)	0.22 (0.41)	0.26 (0.44)	0.21 (0.41)	0.28 (0.45)	0.23 (0.42)	0.23 (0.42)	0.22 (0.42)	0.24 (0.43)
>200	0.32 (0.47)	0.36 (0.48)	0.28 (0.45)	0.28 (0.45)	0.22 (0.42)	0.17 (0.38)	0.40 (0.49)	0.40 (0.49)	0.26 (0.44)	0.40 (0.49)	0.37 (0.48)	0.32 (0.47)	0.33 (0.47)	0.37 (0.48)
Parents' education														
No Second.	0.06 (0.24)	0.01 (0.11)	0.07 (0.25)	0.07 (0.26)	0.14 (0.34)	0.15 (0.36)	0.06 (0.23)	0.13 (0.33)	0.26 (0.44)	0.03 (0.18)	0.07 (0.26)	0.06 (0.24)	0.03 (0.16)	0.06 (0.24)
Upper Second.	---	0.29 (0.46)	0.41 (0.49)	0.30 (0.46)	0.22 (0.41)	0.05 (0.22)	0.29 (0.46)	0.24 (0.42)	0.12 (0.33)	0.03 (0.18)	0.15 (0.35)	0.26 (0.44)	0.03 (0.16)	0.08 (0.28)
Post-Second.	0.27 (0.44)	0.45 (0.50)	0.11 (0.31)	0.09 (0.28)	0.15 (0.36)	0.24 (0.43)	0.36 (0.48)	0.21 (0.41)	0.36 (0.48)	0.48 (0.50)	0.22 (0.41)	0.08 (0.28)	0.50 (0.50)	0.24 (0.43)
University	0.66 (0.47)	0.25 (0.43)	0.40 (0.49)	0.52 (0.50)	0.46 (0.50)	0.43 (0.50)	0.28 (0.45)	0.37 (0.48)	0.21 (0.41)	0.45 (0.50)	0.54 (0.50)	0.57 (0.49)	0.45 (0.50)	0.60 (0.49)
School location														
City	---	0.30 (0.46)	0.24 (0.43)	0.35 (0.48)	0.19 (0.39)	0.33 (0.47)	0.42 (0.49)	---	0.29 (0.46)	0.30 (0.46)	0.11 (0.32)	0.46 (0.50)	0.37 (0.48)	0.17 (0.38)
Town	---	0.61 (0.49)	0.69 (0.46)	0.56 (0.50)	0.74 (0.44)	0.63 (0.48)	0.56 (0.50)	---	0.69 (0.46)	0.47 (0.50)	0.49 (0.50)	0.42 (0.49)	0.32 (0.47)	0.60 (0.49)
Rural	---	0.09 (0.28)	0.07 (0.26)	0.09 (0.28)	0.07 (0.25)	0.05 (0.21)	0.02 (0.15)	---	0.02 (0.14)	0.23 (0.42)	0.40 (0.49)	0.13 (0.33)	0.31 (0.46)	0.22 (0.42)
Parents' attitude														
High	0.15 (0.36)	0.09 (0.29)	0.09 (0.29)	0.13 (0.33)	0.20 (0.40)	0.15 (0.36)	0.23 (0.42)	0.17 (0.38)	0.27 (0.44)	0.22 (0.41)	0.10 (0.29)	0.16 (0.36)	0.23 (0.42)	0.09 (0.29)
Medium	0.70 (0.46)	0.69 (0.46)	0.70 (0.46)	0.73 (0.44)	0.69 (0.46)	0.74 (0.44)	0.67 (0.47)	0.67 (0.47)	0.63 (0.48)	0.66 (0.47)	0.68 (0.47)	0.70 (0.46)	0.62 (0.49)	0.71 (0.45)
Low	0.15 (0.35)	0.22 (0.41)	0.21 (0.40)	0.14 (0.35)	0.10 (0.30)	0.11 (0.31)	0.09 (0.29)	0.16 (0.36)	0.10 (0.30)	0.12 (0.33)	0.22 (0.42)	0.15 (0.36)	0.15 (0.36)	0.20 (0.40)

Values are weighted by the sampling probability of students. Standard deviations are reported in parentheses. For scores, all observations with imputed values were dropped. For student background variables, only the observations for which values of the respective variable were missing were dropped.

Table A4.4
Missing values in percent

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
PIRLS														
Parents' origin	16.08	12.11	12.03	44.69	10.13	12.63	4.41	15.98	3.60	4.45	9.03	15.34	1.42	9.08
Language	2.42	1.16	3.23	0.35	2.31	1.41	1.31	2.67	0.29	1.13	1.51	1.46	1.20	2.11
Books at home	10.76	3.73	8.26	4.03	6.97	2.91	4.06	5.75	2.46	5.34	1.83	7.81	5.16	2.90
Parents' education	18.30	14.94	34.96	48.62	20.92	16.22	8.97	16.20	4.20	11.55	10.06	17.82	1.54	9.42
School location	9.03	16.47	27.73	26.74	24.69	19.17	40.47	28.47	2.51	28.00	19.78	14.89	34.38	10.87
Parents' attit.	17.46	17.24	14.72	45.62	13.85	14.61	6.43	18.02	13.19	8.23	10.98	17.33	6.99	9.36
PISA														
Parents' origin	0.59	0.56	1.91	1.92	5.41	2.99	2.18	1.69	1.19	1.69	2.61	1.93	1.62	0.80
Language	2.08	0.85	8.57	0.91	3.25	1.07	100.0	0.47	5.27	1.85	6.84	4.49	0.85	2.77
Books at home	0.60	0.29	1.49	1.21	0.82	0.50	0.64	0.16	0.54	0.96	0.76	1.07	0.97	0.28
Parents' educ.	4.74	4.33	18.35	15.36	12.04	6.53	5.53	8.36	4.41	12.21	10.78	26.79	20.04	10.06
School location	99.77	0.00	9.66	7.62	10.53	1.62	0.52	100.0	0.00	16.32	4.09	0.00	0.00	1.79
Parents' attit.	0.89	0.33	1.57	1.19	0.93	0.96	0.31	0.56	0.46	0.68	0.64	0.74	2.75	0.46

Table A4.5
Correlation of variables between PIRLS and PISA

	Including imputed values	Without imputed values
Parents' origin	0.84	0.84
Language	0.13	0.04
<i>Books at home</i>		
< 11	0.73	0.83
11-25	0.65	0.79
26-100	0.52	0.77
101-200	0.80	0.79
>200	0.80	0.80
<i>Parents' education</i>		
No Second.	0.43	0.67
Upper Second.	0.05	0.20
Post-Second.	0.16	0.19
University	0.53	0.72
<i>School location</i>		
City	0.42	0.48
Town	0.59	0.56
Rural	0.12	0.32

Correlation coefficients of country means.

Table A4.6

Estimates of student background effects for grade 4 (PIRLS) without imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	17.14 * (2.14)	10.47 * (2.81)	8.81 * (2.02)	18.82 * (5.88)	7.39 † (3.07)	17.40 * (3.60)	15.40 * (2.25)	20.47 * (2.95)	6.91 * (2.30)	20.02 * (2.63)	18.53 * (3.87)	30.40 * (4.62)	12.54 * (2.19)	21.78 * (2.39)
Age	-4.44 (3.37)	-1.36 (3.34)	-8.18 * (2.70)	45.03 * (8.71)	-21.49 * (3.10)	1.59 (4.69)	-9.46 * (3.01)	25.63 * (4.45)	12.95 * (3.74)	-8.67 * (3.18)	25.28 * (6.51)	12.44 † (7.04)	-5.42 (3.66)	4.44 (3.52)
Parents' origin	-4.59 † (2.49)	-7.02 (4.51)	-15.77 * (2.94)	4.92 (5.20)	-6.72 † (3.65)	-11.35 * (4.11)	-6.36 (5.09)	-9.51 † (4.58)	-16.73 * (4.41)	-0.59 (3.01)	2.80 (5.10)	4.85 (4.72)	-16.51 * (6.28)	-9.41 * (2.73)
Language	-25.52 * (3.66)	-23.58 * (6.10)	-25.28 * (4.17)	-49.44 * (8.59)	-21.37 * (4.73)	-15.85 † (6.47)	-42.67 * (6.46)	-28.56 * (4.78)	-29.27 * (7.38)	-12.35 † (6.50)	-34.61 * (6.58)	-44.21 * (8.21)	-36.75 * (11.35)	-27.36 * (4.57)
<i>Books at home</i>														
26-100	6.68 † (3.82)	15.48 * (4.92)	13.38 * (3.70)	7.86 (8.99)	8.39 † (4.20)	17.76 † (4.43)	16.75 * (4.94)	3.36 (6.87)	14.27 * (3.08)	5.20 (5.09)	20.69 † (9.24)	9.20 (5.77)	4.91 (6.45)	12.17 * (3.61)
101-200	10.86 † (4.49)	27.71 * (5.36)	31.26 * (5.67)	31.39 * (10.51)	21.74 * (5.62)	30.15 * (6.07)	20.16 * (5.31)	15.45 † (6.31)	22.90 * (3.98)	10.62 † (5.29)	35.49 * (9.22)	14.15 (8.59)	13.03 † (6.35)	15.07 * (4.33)
>200	18.93 * (4.37)	36.00 * (5.18)	38.63 * (4.23)	28.71 * (9.69)	29.10 * (5.70)	39.81 * (5.40)	37.08 * (5.19)	24.44 * (6.27)	38.47 * (3.93)	14.99 * (5.11)	41.17 * (9.29)	24.78 * (7.48)	16.88 † (6.95)	25.14 * (4.34)
<i>Parents' education</i>														
Upper Second.	11.14 † (6.06)	16.56 † (7.51)	27.49 * (3.80)	6.73 (6.33)	15.40 * (3.60)	12.99 * (4.92)	24.13 * (4.92)	25.08 * (4.46)	19.58 * (3.30)	18.78 * (6.90)	20.27 † (11.46)	18.19 † (10.59)	6.49 (8.19)	16.60 * (4.97)
Post-Second.	24.86 * (6.35)	31.10 * (7.88)	21.70 * (3.91)	26.94 † (15.90)	26.35 * (4.30)	22.46 * (5.81)	25.80 * (7.09)	33.64 * (7.14)	10.02 (6.60)	18.90 * (7.05)	-	35.33 * (10.15)	13.45 † (8.06)	23.72 * (5.39)
University	48.52 * (6.39)	37.18 * (8.08)	35.50 * (4.30)	34.35 * (7.40)	37.51 * (4.39)	46.19 * (6.11)	53.78 * (4.97)	54.20 * (5.15)	27.42 * (3.72)	42.82 * (7.77)	45.44 * (11.07)	59.47 * (10.93)	29.52 * (8.69)	39.81 * (5.33)
<i>School location</i>														
City	2.10 (4.36)	11.20 † (5.31)	-5.53 (4.17)	5.12 (8.72)	5.46 (4.90)	9.28 (5.78)	2.97 (3.73)	3.67 (3.77)	-4.54 (5.29)	8.31 † (4.91)	15.39 † (7.02)	21.71 * (6.68)	-0.10 (6.70)	2.59 (4.90)
Rural	-11.08 † (5.23)	-2.30 (7.62)	3.88 (10.19)	-4.77 (8.29)	8.68 (5.99)	-3.31 (17.53)	-5.83 (5.05)	-1.20 (7.05)	-26.54 † (12.38)	-5.56 (8.36)	7.17 (5.66)	-7.77 (12.28)	-	-1.15 (7.43)
<i>Parents' attitude</i>														
High	12.86 * (2.34)	10.17 * (3.26)	12.46 * (2.85)	25.86 * (6.55)	11.69 * (3.39)	9.76 † (3.80)	9.66 * (3.04)	-0.78 (3.21)	18.56 * (2.60)	9.22 * (2.04)	12.76 † (5.00)	32.99 * (5.23)	11.29 * (3.20)	9.40 * (2.44)
Low	-10.07 † (5.37)	-1.33 (8.04)	0.28 (3.49)	-4.20 (11.69)	-2.48 (8.15)	7.98 (7.65)	6.70 (7.97)	-6.12 (9.14)	-2.12 (4.87)	-15.04 † (7.61)	2.16 (9.31)	-2.46 (11.78)	3.42 (6.19)	-8.44 (5.36)
Observations	5,288	1,948	3,066	1,095	1,783	1,610	2,376	1,954	2,838	1,740	2,297	1,537	2,315	4,631
R-Square	0.16	0.16	0.28	0.21	0.27	0.24	0.28	0.17	0.17	0.20	0.16	0.25	0.15	0.18

Coefficients from cluster-robust linear regressions (CRLR), weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Schools are clusters.

Significance levels: * 1 percent, † 5 percent, ‡ 10 percent. Dependent variable: PIRLS original reading literacy score. All observations with imputed values have been dropped.

Table A4.7
Estimates of student background effects for grade 4 (PIRLS) from SFE model

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	17.38* (1.59)	10.06* (2.55)	10.30* (1.49)	20.31* (2.49)	8.55* (2.18)	17.02* (2.33)	13.19* (1.80)	18.22* (2.25)	7.08* (2.04)	18.62* (2.20)	22.98* (3.12)	26.52* (3.62)	10.21* (1.42)	22.22* (2.15)
Age	-6.40† (2.63)	-4.25‡ (2.46)	-13.83* (1.70)	32.93* (4.71)	-26.25* (2.10)	-0.74 (3.85)	-14.36* (2.21)	23.07* (3.98)	6.00‡ (3.20)	-10.79* (2.25)	27.03* (4.86)	3.14 (5.38)	-5.19† (2.12)	2.98 (3.19)
Parents' origin	0.34 (2.06)	-6.46‡ (3.33)	-11.65* (1.95)	1.69 (3.32)	-2.23 (2.46)	-16.67* (3.60)	-12.99* (3.69)	-11.41* (3.66)	-15.59* (3.80)	-1.95 (2.57)	-1.71 (4.22)	0.08 (4.12)	-6.53* (1.77)	-5.18† (2.30)
Language	-22.64* (2.79)	-21.75* (4.61)	-30.30* (2.51)	-37.44* (4.43)	-16.89* (3.04)	-9.23 (5.65)	-33.22* (4.26)	-26.83* (3.62)	-20.34* (6.10)	-18.36* (4.34)	-29.39* (5.59)	-33.29* (5.25)	-16.45* (3.52)	-22.00* (3.29)
Books at home														
26-100	4.99† (2.89)	11.91* (4.26)	12.51* (2.13)	-3.01 (4.51)	12.78* (2.29)	17.91* (3.41)	15.68* (3.59)	11.63† (5.43)	9.10* (2.71)	-3.09 (6.65)	20.21* (7.11)	10.24† (4.68)	9.60* (2.40)	15.06* (3.28)
101-200	14.53* (3.24)	27.44* (4.67)	26.87* (2.86)	12.69† (5.43)	20.00* (3.82)	29.40* (5.40)	20.24* (3.70)	25.46* (5.52)	22.43* (3.50)	3.57 (6.53)	30.91* (7.64)	21.62* (6.52)	12.34* (2.46)	18.36* (3.98)
>200	21.51* (3.41)	32.57* (4.26)	32.99* (2.79)	23.67* (5.11)	31.58* (3.57)	36.74* (4.22)	31.50* (3.59)	22.21* (5.43)	34.52* (3.10)	4.82 (7.49)	38.47* (6.83)	31.61* (5.51)	14.88* (2.60)	23.19* (3.75)
Parents' education														
UpperSecond.	7.53 (4.73)	21.39* (8.03)	24.23* (2.95)	10.05‡ (5.47)	10.51* (2.95)	11.56* (4.16)	20.58* (3.75)	19.30* (3.79)	17.53* (2.50)	21.67* (5.56)	32.94* (8.41)	17.80† (8.43)	3.34 (4.86)	13.46* (4.43)
Post-Second.	17.20* (5.00)	38.76* (8.39)	17.45* (3.07)	15.45 (11.09)	22.27* (3.37)	30.56* (5.48)	22.59* (5.93)	29.70* (5.80)	15.30† (7.00)	27.81* (6.10)		33.77* (8.30)	11.89† (4.88)	20.87* (4.84)
University	40.68* (4.86)	34.35* (8.57)	31.14* (3.39)	34.79* (4.64)	26.72* (3.31)	46.92* (5.39)	41.95* (4.51)	36.66* (4.34)	24.18* (3.16)	38.65* (6.43)	50.92* (8.70)	42.32* (9.14)	19.31* (5.05)	27.67* (4.78)
R-Squared	0.32	0.29	0.34	0.34	0.37	0.42	0.42	0.21	0.34	0.36	0.25	0.40	0.56	0.25

Coefficients from school-fixed effects (SFE) estimations, weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Significance levels: * 1 percent, † 5 percent, ‡ 10 percent. Dependent variable: PIRLS original reading literacy score.

Table A4.8

Estimates of student background effects for grade 9 (PISA) without imputed values

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	26.74* (1.50)	20.60* (3.49)	20.44* (3.22)	20.87* (4.66)	15.39* (2.57)	33.15* (3.71)	23.34* (3.50)	34.58* (3.17)	25.29* (4.25)	43.90* (4.36)	33.66* (3.16)	35.55* (4.81)	30.00* (2.60)	27.92* (2.64)
Age	-0.96 (2.83)	-29.27* (5.44)	-49.51* (6.75)	16.75* (5.62)	0.69 (3.95)	8.42 (5.48)	-35.30* (5.65)	9.37‡ (5.08)	11.63* (4.25)	-22.65* (8.68)	20.52* (5.29)	27.62* (6.11)	-10.43‡ (5.57)	14.53* (4.40)
Parents' origin	2.46 (2.27)	7.61† (3.58)	-5.39 (5.72)	6.27 (4.99)	-8.81* (3.01)	4.91 (6.12)	-3.27 (6.02)	-7.41 (7.23)	10.94† (4.49)	-10.14‡ (5.71)	-3.22 (5.67)	1.53 (4.33)	6.19 (3.94)	-3.84 (3.74)
Language	-28.98* (3.08)	16.10 (16.51)	-42.94* (9.99)	-33.31* (8.13)	-21.93* (4.76)	-27.11* (9.42)	-	-31.57† (13.27)	-27.28* (5.80)	-8.92 (13.01)	-49.01* (8.86)	-73.12* (8.43)	-7.08 (13.25)	-39.79* (8.17)

Books at home

26-100	19.27* (2.18)	24.06* (3.51)	26.37* (4.21)	24.64* (4.23)	17.19* (2.99)	18.17* (3.33)	24.35* (3.79)	16.07* (5.75)	15.70* (4.22)	1.24 (8.79)	18.64* (6.56)	20.96* (5.35)	14.74* (3.75)	25.51* (4.14)
101-200	28.91* (2.30)	41.90* (3.76)	41.47* (4.55)	50.67* (4.81)	28.78* (3.35)	28.58* (4.16)	48.11* (4.11)	35.98* (4.99)	28.00* (4.81)	24.87* (7.05)	47.44* (6.24)	42.37* (5.43)	42.46* (4.32)	36.60* (4.07)
>200	39.35* (2.45)	61.91* (4.21)	61.59* (5.14)	68.86* (5.03)	40.31* (3.57)	42.94* (6.20)	69.11* (4.65)	50.57* (4.73)	36.20* (5.64)	42.01* (7.60)	56.80* (6.31)	54.57* (5.59)	51.02* (4.70)	60.20* (4.00)

Parents' education

Upper Second.	-	1.89 (9.87)	28.43* (6.17)	18.80* (5.48)	3.01 (3.65)	17.53* (6.52)	25.91* (5.50)	18.39* (5.06)	12.99* (4.15)	13.70 (13.78)	16.15† (7.12)	29.18* (7.37)	23.95† (10.35)	9.10 (6.92)
Post-Second.	-	25.52* (9.78)	37.58* (7.07)	53.63* (7.29)	9.44† (3.89)	28.25* (4.15)	51.39* (6.06)	21.57* (4.61)	22.41* (3.66)	24.55* (8.91)	8.89 (7.25)	44.28* (8.15)	37.48* (7.05)	13.75† (5.55)
University	21.51* (1.64)	56.90* (10.20)	52.00* (6.13)	35.44* (5.66)	8.85† (3.55)	39.71* (4.78)	75.20* (6.92)	30.30* (4.91)	27.08* (4.08)	31.75* (9.6)	19.35* (7.23)	47.57* (7.35)	45.16* (7.94)	11.95† (5.77)

School location

City	-	-0.33 (6.53)	-5.73 (7.97)	0.72 (6.80)	5.59 (5.31)	13.49 (8.68)	16.78† (6.86)	-	5.21 (8.63)	17.59‡ (9.21)	3.18 (6.24)	12.76† (5.50)	17.85† (7.61)	5.79 (5.04)
Rural	-	-20.05* (7.61)	-33.49* (10.67)	-7.76 (7.29)	-0.279 (6.95)	-13.24 (12.64)	-28.44‡ (16.85)	-	6.44 (15.11)	-32.90* (8.61)	-9.76‡ (5.20)	-14.55† (7.03)	-17.54† (8.91)	-6.93‡ (3.85)

Parents' attitude

High	21.91* (2.20)	9.38† (3.79)	10.98† (5.34)	29.47* (4.67)	7.35* (2.26)	6.01 (4.15)	7.83* (2.67)	13.93* (4.12)	1.59 (3.28)	12.59* (4.51)	23.19* (4.92)	20.76* (5.06)	8.45* (2.97)	14.84* (4.55)
Low	-26.01* (1.94)	-20.71* (2.61)	-21.91* (3.44)	-26.31* (4.18)	-11.28* (3.23)	-19.98* (4.47)	-11.15* (4.15)	-27.73* (4.39)	-21.94* (4.48)	-15.69* (5.66)	-33.91* (3.94)	-19.94* (4.30)	-18.37* (3.21)	-32.23* (3.36)
Observations	26,815	4,865	3,398	7,007	3,353	4,036	4,401	2,868	4,283	2,711	3,160	2,526	5,087	3,728
R-Square	0.20	0.30	0.37	0.22	0.50	0.23	0.42	0.16	0.26	0.29	0.20	0.22	0.22	0.22

Coefficients from cluster-robust linear regressions (CRLR), weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Schools are clusters.

Significance levels: * 1 percent, † 5 percent, ‡ 10 percent. Dependent variable: PIRLS original reading literacy score. All observations with imputed values have been dropped.

Table A4.9
Estimates of student background effects for grade 9 (PISA) from SFE model

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female	27.29* (1.38)	15.65* (2.12)	12.53* (2.14)	22.63* (2.65)	14.92* (1.74)	21.17* (2.33)	11.88* (2.05)	35.68* (3.05)	11.45* (2.42)	35.49* (3.09)	36.50* (3.44)	45.70* (3.84)	23.26* (2.02)	30.45* (2.61)
Age	-5.81† (2.47)	-27.49* (4.52)	-23.48* (3.91)	17.65* (4.22)	1.47 (3.18)	10.45† (4.00)	-17.13* (4.05)	10.73† (5.32)	4.30 (3.09)	-23.80* (6.94)	19.79* (5.26)	31.25* (5.08)	-16.78* (3.70)	15.87* (4.22)
Parents' origin	5.11† (2.04)	6.99† (3.01)	-11.74* (3.57)	-3.56 (4.31)	-7.93* (2.51)	8.07 (5.07)	-6.07 (4.08)	-9.23 (7.26)	3.72 (3.98)	0.49 (4.27)	-9.73† (4.68)	9.25* (3.43)	2.89 (2.72)	2.70 (3.46)
Language	-29.17* (2.50)	-9.87 (7.74)	-27.76* (4.69)	-25.18* (5.75)	-15.98* (3.76)	-28.69* (8.87)	-	-25.64† (11.91)	-10.41* (2.24)	-27.07* (9.12)	-26.82* (5.00)	-71.52* (4.98)	-13.88† (5.75)	-32.39* (5.22)
<i>Books at home</i>														
26-100	20.02* (1.90)	16.80* (3.12)	13.54* (3.33)	23.61* (3.67)	16.10* (2.41)	11.96* (2.55)	11.94* (2.86)	12.02† (5.58)	8.28* (2.77)	2.93 (5.52)	17.24* (5.51)	25.25* (4.35)	8.62* (2.71)	22.81* (4.08)
101-200	31.79* (2.08)	27.35* (3.16)	18.56* (3.42)	41.22* (3.28)	24.74* (2.72)	17.05* (3.00)	20.48* (2.65)	30.93* (4.78)	11.29* (2.65)	23.40* (5.25)	48.06* (4.92)	45.04* (4.19)	28.54* (2.82)	39.26* (3.69)
>200	41.92* (1.86)	40.90* (3.38)	22.95* (3.60)	47.87* (3.90)	31.69* (2.96)	21.11* (3.87)	27.77* (2.87)	43.77* (4.73)	9.54* (3.26)	29.21* (5.50)	53.16* (4.87)	55.03* (4.73)	27.30* (2.63)	59.68* (3.51)
<i>Parents' education</i>														
Upper Second.	-	6.54 (9.06)	15.39* (3.57)	3.44 (4.90)	6.53† (2.95)	11.07† (4.29)	8.15† (3.79)	14.79* (5.03)	3.79 (2.96)	22.51† (10.72)	24.05* (6.35)	19.06* (6.56)	20.76† (8.26)	19.69* (6.09)
Post- Second.	18.39* (2.79)	16.16† (9.03)	16.09* (4.64)	29.19* (6.48)	7.86† (3.16)	13.28* (2.93)	11.75* (3.91)	19.84* (4.56)	6.06* (2.27)	22.03* (7.93)	10.75† (5.99)	31.60* (7.70)	21.49* (6.12)	21.93* (5.00)
University	28.33* (2.59)	29.21* (9.28)	18.47* (4.29)	11.96† (5.10)	2.32 (2.75)	14.19* (2.95)	11.93* (4.06)	22.61* (4.61)	-0.65 (2.77)	25.76* (8.24)	16.70* (5.58)	24.71* (6.65)	22.68* (6.50)	12.45† (5.12)
R-Squared	0.35	0.56	0.74	0.35	0.61	0.56	0.72	0.22	0.61	0.47	0.26	0.35	0.51	0.27

Coefficients from school-fixed effects (SFE) estimations, weighted by students' sampling probabilities. Cluster-robust standard errors in parentheses. Dummies for grade levels are included in regressions. Significance levels: * 1 percent, † 5 percent, ‡ 10 percent. Dependent variable: PISA original reading literacy score.

Table A4.10

Changes in the absolute size of effects between grade 4 and grade 9 using rescaled scores

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Female								+	+	+			+	
Age	-	+	+	-	-		+	-		+	-		-	+
Parents' origin		-	+						-	+			-	-
Language	-		+			+					-		-	
<i>Books at home</i>														
26-100	+		+	-								-		
101-200				+									+	+
>200			+	+					-	+		+	+	+
<i>Parents' education</i>														
Upper Second.			-	-				-	-		-			
Post-Second.					-	-		-		-			+	-
University	-			-	-	-		-	-	-	-	-		-
<i>School location</i>														
City		-					+							
Rural		+							-	+	+			
<i>Parents' attitude</i>														
High	+					-		+	-			-		
Low	+		+	+		+			+		+		+	+

A + indicates an increase, a - a decrease in the absolute size of the effect between PIRLS and PISA. All effects are significant at the 10 percent-significance-level. Results are based on rescaled scores.

Table A4.11
Cross-sectional results from PIRLS

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	.04* (.01)	.79* (.18)		
One parent univ. degree	.02* (.01)	.84* (.18)		
<i>Parents' attitude</i>				
High	.05* (.008)			
Low	.03‡ (.02)			

Interaction effects between institutions and student background effects are presented. Canada and England are included in the regressions based on equation (4.4).
Dependent variable: PIRLS original reading literacy score.
FGLS estimations with heteroskedastic panels.
Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.12
Cross-sectional results from PIRLS from SFE model

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	.05* (.01)	.79* (.18)		
One parent univ. degree	.02† (.01)	.89* (.18)		
<i>Parents' attitude</i>				
High	.05* (.008)			
Low	.04† (.02)			

Interaction effects between institutions and student background effects are presented. Canada and England are included in the regressions based on equation (4.4).
Dependent variable: PIRLS original reading literacy score.
FGLS estimations with heteroskedastic panels and school-fixed effects.
Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.13
Cross-sectional results from PISA

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.08* (.01)	-.45* (.13)	4.20* (.80)	
One parent univ. degree	.02‡ (.01)	-.89* (.12)	.72 (.80)	
<i>Parents' attitude</i>				
High	-.03* (.01)			3.23* (.52)
Low	.05* (.01)			-1.28† (.57)

Interaction effects between institutions and student background effects are presented.
Canada and England are included in the regressions based on equation (4.4).
Dependent variable: PISA original reading literacy score.
FGLS estimations with heteroskedastic panels.
Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.14

Cross-sectional results from PISA from SFE model

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.08* (.01)	-.51* (.14)	-3.23* (.87)	
One parent univ. degree	.01 (.01)	-.46* (.13)	-1.29 (.87)	
<i>Parents' attitude</i>				
High	-.03† (.01)			2.65* (.57)
Low	.06* (.01)			-1.88* (.63)

Interaction effects between institutions and student background effects are presented.

Canada and England are included in the regressions based on equation (4.4).

Dependent variable: PISA original reading literacy score.

FGLS estimations with heteroskedastic panels and school-fixed effects.

Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.15
 Test on equal coefficients between PIRLS and PISA

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	8.62*	5.64†		
One parent univ. degree	0.40	8.05*		
<i>Parents' attitude</i>				
High	6.24†			
Low	0.86			

Chi2 statistic for Wald test on equal coefficients between the results in Tables A4.11 and A4.13.
 Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.16
 Test on equal coefficients between PIRLS and PISA from SFE model

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	8.10*	5.74†		
One parent univ. degree	0.50	6.09†		
<i>Parents' attitude</i>				
High	5.41†			
Low	-1.29			

Chi2 statistic for Wald test on equal coefficients between the results in Tables A4.12 and A4.14.
 Significance-level: * 1 percent. † 5 percent. ‡ 10 percent.

Table A4.17
Results from two-step model

	Instruction time	Share of students in private schools	School types	School autonomy
>200 Books at home	-.0034† (.0015)	4.64† (1.90)	9.24 (5.46)	
One parent univ. deg.	.0014 (.0015)	-1.02 (1.99)	11.51‡ (5.71)	
<i>Parents' attitude</i>				
High	.0007 (.001)			3.21‡ (1.76)
Low	.0003 (.0007)			1.26 (.99)

Seemingly unrelated regressions based on equation (4.7), using small-sample t-statistics. Similar to the estimates presented in Table 4.5, separate regressions were performed for those institutional variables with missing values, always including all possible interaction effects.

Dependent variables: Difference in student background effects between grade 4 and grade 9.

Coefficients significant at the 10 percent-level are printed in bold. Standard errors in parentheses.

Number of observations between ten and 14.

Table A4.18
Institutional variables

	CAN	CZE	DEU	ENG	FRA	GRC	HUN	ISL	ITA	LVA	NOR	NZL	RUS	SWE
Institutions														
Number of school types in primary education ¹	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Number of school types in lower secondary ¹	1	2	3	3	1	2	2	1	3	1	1	2	2	1
Yearly instruction time in minutes in PIRLS ²	24,583	24,741	23,560	23,462	23,563	18,152	—	19,614	29,289	14,145	19,742	29,962	16,623	24,243
Yearly instruction time in minutes in PISA ³	58,576	56,804	54,547	57,369	61,304	47,947	52,231	50,935	61,769	51,142	—	57,881	51,534	53,615
Share of students in private schools in primary education in 2000 (in percent) ⁴	6.5	.9	2.2	4.7	15	7	5.1	1.4	6.4	2	1.5	2	0.4	4
Share of students in private schools in lower secondary education in 2000 (in percent) ⁴	8	2	7	7	21	5	5	1	4	4	2	4	1	3
Percentage of women among teaching staff in public and private institutions in primary education, based on head counts ⁵	68	84.4	81.2	81.1	80	—	85	77.8	94.8	—	—	83.8	98.7	80.4
Percentage of women among teaching staff in public and private institutions at lower secondary, based on head counts ⁵	68	82	59	59	64	64	85	77	73	88	72	65	89	62
Index of school autonomy in lower secondary ⁶	5.68	7.77	3.90	6.47	—	3.10	6.89	6.46	2.43	—	5.12	6.67	3.60	6.05
Educational expenditure														
Expenditure on educational institutions as a percentage of GDP 2000 ⁷	6.4	4.6	5.3	5.3	6.1	4.0	5.0	6.3	4.9	—	5.9	5.8	—	6.5
Private education expenditure as a percentage of total education expenditure (all levels), 1998 ⁸	11.0	12.8	21.6	5.7	5.8	—	11.6	11.0	3.9	—	1.9	2.6	—	2.7
(1995)														
Country and population														
Female employment rate in % ⁹	74.0	73.7	71.1	73.1	69.6	52.6	61.7	87.4	50.7	—	81.5	70.6	—	81.7
Stock of foreign population in % (1999) ¹⁰	17.411	2.2	8.8	3.8	5.6	2.7	1.3	1.5 ¹²	2.2	41.8 ¹³	4.0	24.2 ¹³	—	5.5
Urban population in % (2003) ¹⁴	80.4	74.3	88.1	89.1	76.3	60.8	65.1	92.8	67.4	66.2	78.6	85.9	73.3	83.4

1 Source: National descriptions in Eurybase; Statistics Canada, New Zealand. The number of school types in England has alternatively been set to 1, which does not affect the results.

2 Source: PIRLS 2001 database.

3 Source: OECD Education at a Glance 2002, C2.4.

4 Source: OECD Education at a Glance 2003, Tab. B2.1a.

5 Source: EdStats, <http://devdata.worldbank.org/edstats/ThematicDataOnEducation/PrivateEducationExpenditure/tab21.xls>.

6 Source: EdStats, <http://www.oecd.org/dataoecd/23/41/2508596.pdf>.

7 Source: <http://www.oecd.org/dataoecd/10/35/2079451.pdf>.

8 Source: Berliner Institut für Vergleichende Sozialforschung, http://www.ernz-berlin.de/Statistik_2/lat/lat_01.htm.

9 Source: UN Department of Economic and Social Affairs, Population Division, "Urban and Rural Areas 2003".

Figure A4.1
Mean scores and standard deviations for PIRLS (slope not significant)

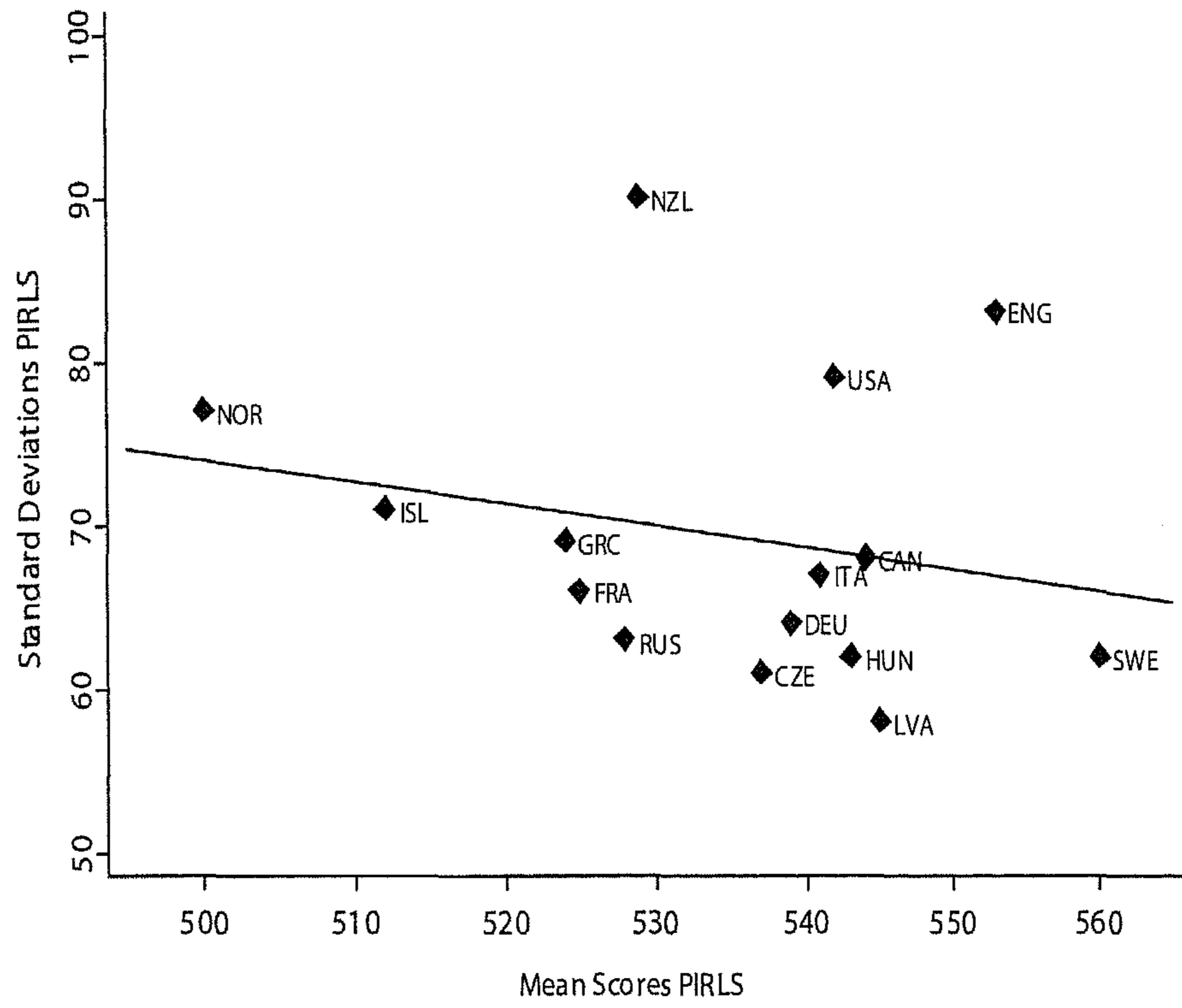
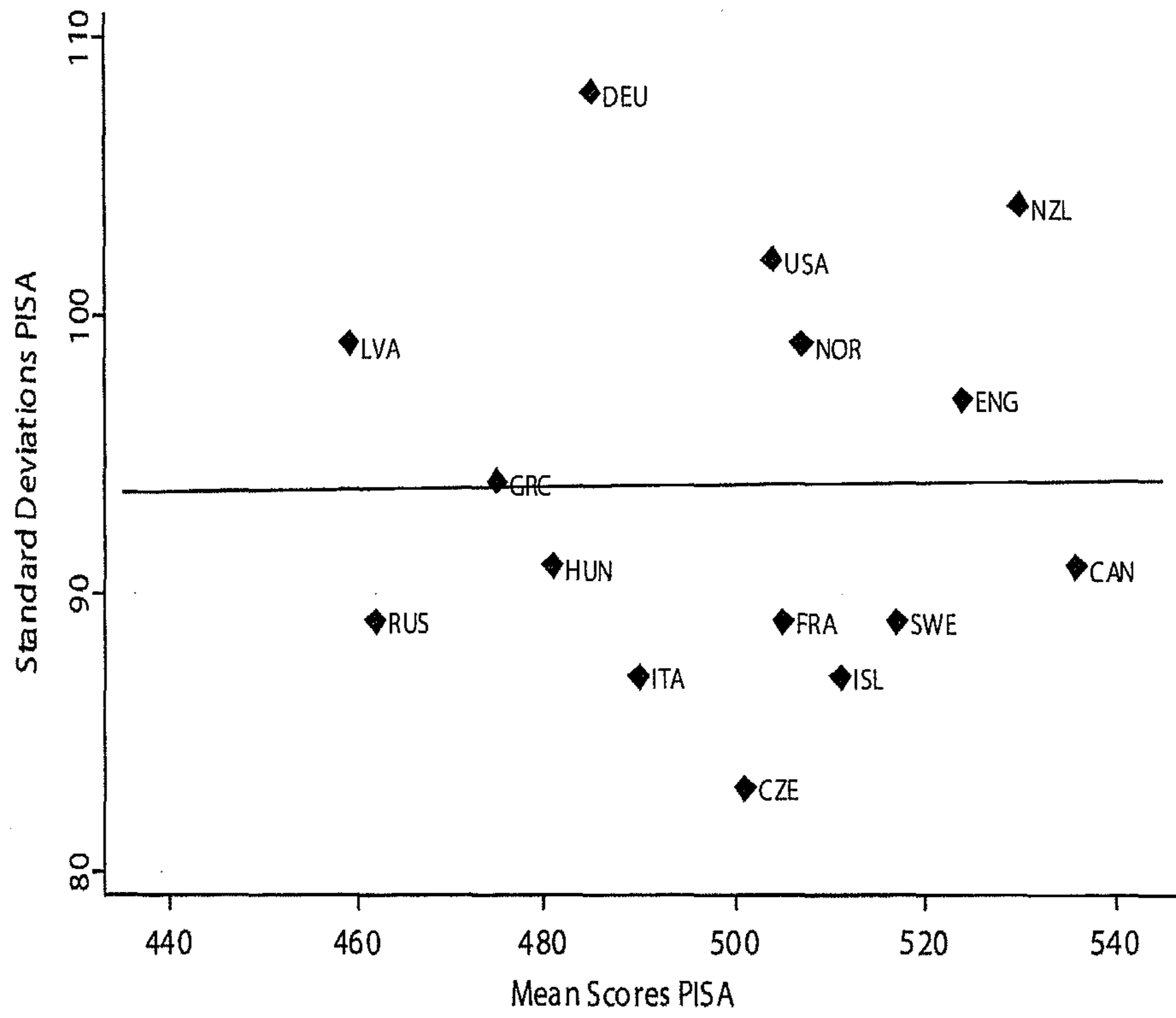


Figure A4.2
Mean scores and standard deviations for PISA (slope not significant)



Summary in Dutch

In dit proefschrift wordt onderzocht wat de productie is van het onderwijs op de leerplichtige leeftijd, en dan met name de prestaties van scholieren in het lager voortgezet onderwijs. Het onderzoek richt zich vooral op de theoretische en empirische analyse van het verband tussen de sociale achtergrond van scholieren en de kwaliteit van de school, inclusief de onderwijsinstituties, en de prestaties van scholieren. Naast het analyseren van de relatie tussen de inputfactoren van het onderwijssysteem en de prestaties van scholieren, wordt in dit onderzoek getracht een verklaring te vinden voor de verschillen in gelijkheid van onderwijskansen per onderwijsinstelling.

Voorafgaand aan de empirische analyse wordt een overzicht gegeven van de relevante literatuur en een beschrijving van het methodologische kader voor de latere onderdelen van het onderzoek. Allereerst wordt een theoretisch model voor de onderwijsproductie ontwikkeld, waarbij rekening wordt gehouden met de heterogeniteit van de achtergronden van de scholieren. Daarmee kan de interactie tussen de onderwijsinstituties en de effecten van de achtergrond van de scholieren in het model worden geanalyseerd. Het model is gebaseerd op een multiplicatieve Cobb-Douglas productiefunctie. Het evenwicht wordt bepaald door een gezamenlijke maximalisering van scholieren en ouders over de inspanningen van de scholieren respectievelijk de betrokkenheid van de ouders. Het model wordt vervolgens toegepast om te bepalen wat de effecten zijn van de institutionele aspecten aanwezigheid van verschillende schooltypen, onderwijstijd, omvang van de particuliere schoolsector en de autonomie van scholen op de prestaties van scholieren met een hoge en lage sociale achtergrond, oftewel de mate van gelijke onderwijskansen. Zowel het aantal verschillende schooltypen als de omvang van de particuliere schoolsector zijn van invloed op de mate van differentiatie in het onderwijssysteem. Grotere differentiatie biedt scholieren meer keuzemogelijkheden. Aan de andere kant geeft het theoretische model aan dat wanneer scholieren met een hogere sociale achtergrond minder beperkingen en daardoor meer keuze hebben ten aanzien van de school waar zij heen gaan, de aanwezigheid van een groot aantal verschillende schooltypen en een grote particuliere schoolsector zorgen voor minder onderwijskansen. Een soortgelijke redenering geldt voor het effect van de autonomie van scholen. Ouders kunnen makkelijker invloed uitoefenen op beslissingen wanneer deze worden genomen op lokaal niveau dan op centraal niveau. Wanneer ouders van scholieren met een hogere sociale achtergrond meer betrokkenheid tonen bij het onderwijs van hun kind, dan is het waarschijnlijker dat zij invloed zullen uitoefenen op de betreffende school ten behoeve van hun kind. Het laatste institutionele aspect waarvan het effect door het theoretische model wordt bekeken, is de hoeveelheid onderwijstijd. Naarmate de hoeveelheid onderwijstijd toeneemt, neemt het belang van thuis af bij het aanleren van vaardigheden. Derhalve speelt de achtergrond van scholieren een minder belangrijke rol in hun prestaties en nemen de onderwijskansen toe wanneer scholieren meer tijd samen op school doorbrengen. De resultaten van de theoreti-

sche analyse bieden inzicht in welke effecten van bepaalde instituties op de onderwijskansen te verwachten zijn en op welke manier deze effecten worden veroorzaakt.

Daarna worden de empirische aspecten van de opbrengst van onderwijs besproken. In het onderzoek worden verschillende meetwaarden voor het bereikte onderwijsniveau gepresenteerd en wordt het nut daarvan voor economische analyses vergeleken. Bovendien wordt ingegaan op de beperkingen van de prestatiegegevens van scholieren en de problemen bij het schatten van de productiefuncties. Ook worden mogelijke oplossingen aangegeven waarmee deze problemen zouden kunnen worden verholpen.

Met de eerste empirische analyse worden de achterliggende redenen onderzocht voor het verschil in prestaties tussen scholieren uit Finland en uit Duitsland in het lager voortgezet onderwijs in de PISA studie van het jaar 2000. Finland is een van de hoogst scorende landen in recente prestatietests en vertoont weinig ongelijkheid in de verdeling van de testcores. Duitse scholieren scoren daarentegen onder het OESO-gemiddelde in PISA en hun testcores lopen sterk uiteen. Bovendien is het systeem van voortgezet onderwijs in Duitsland gebaseerd op verschillende schooltypen, terwijl er in Finland slechts één schooltype is.

Finse scholieren doen het in PISA op vijftienjarige leeftijd veel beter dan Duitse, met name in de laagste regionen van de verdeling van de testcores. Terwijl het verschil tussen de twee landen 30 punten bedraagt voor het deciel best presterende scholieren, loopt het verschil op tot meer dan 75 testcorepunten voor het laagste deciel. Bij het schatten van onderwijsproductiefuncties voor beide landen afzonderlijk, valt te zien dat de prestaties van scholieren in Duitsland in grotere mate worden bepaald door waarneembare kenmerken dan in Finland. Met name het effect van sociale achtergrond is groter in Duitsland. Bij uitsplitsing van het verschil tussen de testcores blijkt dat verschillen in achtergrond bij scholieren en hulpbronnen nauwelijks een verklaring bieden voor welk deel van het verschil in prestaties dan ook. Daarentegen hebben Duitse scholieren gemiddeld een hogere sociale achtergrond dan Finse scholieren. Alleen voor de laagste decielen van de verdeling van scholieren is de achtergrond van de scholieren gunstiger bij de Finse dan bij de Duitse scholieren. Dit gegeven kan een deel van het verschil tussen de scores verklaren. Het gebruik van hulpbronnen en de omzetting van eigenschappen van scholieren in prestaties worden echter bevonden doelmatiger te zijn in Finland dan in Duitsland. Dit kan ook een groot deel van het verschil tussen de testcores verklaren. Het grootste verschil tussen de onderwijsprocessen in de twee landen ligt bij de gelijkheid van onderwijskansen. Deze is veel hoger in Finland dan in Duitsland. Bij de analyse van de twee landen wordt echter uitsluitend gebruik gemaakt van de variatie die waarneembaar is binnen elk land. Verschillen in institutionele aspecten die zich alleen tussen de landen voordoen, zoals de aanwezigheid van verschillende schooltypen, kunnen met deze benadering niet worden geanalyseerd. Hoewel de informatie over scholieren en scholen in de PISA-gegevens zeer uitgebreid is, wordt de analyse beperkt door het feit dat de gegevens een dwarsdoorsnede vormen. Voor het bepalen van de effecten van hulpbronnen bijvoorbeeld moeten sterkere aannames worden gedaan over selectie en eerdere inputs. Daarom worden deze resultaten met voorzichtigheid geïnterpreteerd.

De analyse over meerdere landen heen, die het tweede deel van de empirische analyse uitmaakt, gaat uit van de verschillen in institutionele aspecten van de onderwijssystemen, als verklaring van de grote variatie van onderwijskansen die ook waargenomen werd in het eerste deel van de analyse. Om op consistente wijze het verband tussen instituties en onderwijskansen te kunnen schatten, worden de onderwijssystemen van een steekproef van voornamelijk Europese landen geanalyseerd op twee momenten: het einde van het basisonderwijs en het einde van het lager voortgezet onderwijs. Hierbij wordt gebruik gemaakt van de gegevens op scholierenniveau van respectievelijk PIRLS 2001 en PISA 2000. De veranderingen tussen de twee meetmomenten in effecten van de achtergrond van scholieren worden in wisselwerking gebracht met de veranderingen in de institutionele aspecten aantal aanwezige schooltypen, omvang van de particuliere schoolsector, hoeveelheid onderwijstijd en autonomie van scholen. Daarmee kan grotendeels worden gecontroleerd voor landspecifieke effecten en kunnen de institutionele effecten worden geïdentificeerd. Met de empirische analyse worden de hypothesen over de determinanten van onderwijskansen getest vanuit het theoretische model. De resultaten wijzen uit dat een hoge mate van differentiatie in het onderwijssysteem, die het gevolg is van de aanwezigheid van een groot aantal verschillende schooltypen en een grote particuliere schoolsector, de onderwijskansen voor scholieren met een lagere sociale achtergrond verminderen. Hoe groter het aantal verschillende schooltypen, des te groter dit effect. De hoeveelheid onderwijstijd daarentegen had een positief effect op de mate van gelijkheid van onderwijskansen. Een hogere mate van autonomie van de school heeft een positieve invloed op scholieren van wie de ouders zeer betrokken zijn bij het onderwijs van hun kind in vergelijking met scholieren van wie de ouders minder om onderwijsprestaties geven. Voor de meeste hypothesen uit het theoretische model werd derhalve empirische ondersteuning gevonden.

Andere resultaten zijn de verschuivingen in het belang van de achtergrondfactoren van scholieren voor de onderwijsprestaties naarmate de scholier ouder wordt. De houding van ouders tegenover leren wordt belangrijker terwijl gezinskenmerken zoals de afkomst of het opleidingsniveau van de ouders minder bepalend zijn voor de prestaties van scholieren op latere leeftijd dan in het basisonderwijs. De analyse wordt beperkt door het kleine aantal landen waarvoor gegevens beschikbaar zijn voor zowel het basisonderwijs als het voortgezet onderwijs, en derhalve door de kleine variatie op landniveau. Bovendien steunen de resultaten op de vergelijkbaarheid van twee dwarsdoorsnedenonderzoeken en de kwantitatieve informatie die beschikbaar is over de institutionele aspecten van onderwijssystemen.

Het proefschrift wordt afgesloten met een samenvatting en een synthese van de resultaten. Tot slot worden mogelijke beleidsconclusies en interessante nieuwe richtingen voor het onderzoek vermeld. De grootste uitdaging voor het Duitse onderwijssysteem is het bevorderen van de prestaties van scholieren die het slecht doen. Dit kan worden bewerkstelligd door het verhogen van de onderwijskansen voor achterstandskinderen met een lage sociale achtergrond. Minder differentiatie van het onderwijssysteem bleek bevorderlijk voor gelijke onderwijskansen. Zowel een groot aantal verschillende schooltypen in het lager voortgezet onderwijs als een grote particuliere schoolsector dragen bij aan meer differentiatie en dienen derhalve te worden teruggedrongen indien de prestaties van de kinderen met een sociale

achterstand moeten worden verbeterd. Bovendien moeten beleidsmakers zich ervan bewust zijn dat meer autonomie van scholen ook de invloed van de ouders doet toenemen. Aan de andere kant vermindert meer onderwijstijd de effecten van de gezinsachtergrond op de prestaties. Met name in onderwijssystemen die werken met onderwijs dat zich beperkt tot dagdelen, kunnen de prestaties van slecht presterende scholieren die de grootste sociale achterstand hebben worden verbeterd door de lestijd en begeleiding uit te breiden. De eerstgenoemde beleidslijnen van het afschaffen of uitstellen van de differentiatie naar verschillende schooltypen en het beperken van het opzetten van particuliere scholen zijn grotendeels kostenneutraal, tenminste op de langere termijn, maar voor het uitbreiden van de onderwijstijd en begeleiding zijn extra middelen nodig voor het onderwijssysteem.

Curriculum Vitae

Andreas Ammermüller was born on 11 March 1978 in Bonn – Bad Godesberg (Germany). He obtained the Abitur in 1997 from the Amos-Comenius-Gymnasium in Bonn. After his civil service he started to study economics at the University of Maastricht in 1998 with a specialization in 'International Economic Studies'. In 1999 he received the 'Propaedeuse' with distinction. During the course of his studies, he also followed courses at the Universities of Liege and Aachen as part of the „International Management and Economics Programme” (IMEP) and spent the winter semester 2001/2002 as an Erasmus student at the University of Nice. In September 2002 he finalized his studies with a Master thesis under the supervision of Professor Hans Heijke and received the Master's degree in Economics at the University of Maastricht.

In December 2002 he joined the Centre for European Economic Research (ZEW) in Mannheim as a research fellow. In his research at the department of 'Labour Markets, Human Resources, and Social Policy' his main fields of interest are education and labor economics. In various research projects for the European Commission, German government departments and science foundations he addressed the issues of wage inequality, temporary work, wage subsidies, returns to education and human capital formation. He took courses at the graduate school of economics in Mannheim and visited the Centre for Economic Performance (CEP) at the London School of Economics in 2006.

Roa Dissertation Series

1. Lex Borghans (1993), Educational Choice and Labour Market Information, Maastricht: Research Centre for Education and the Labour Market (ROA).
2. Frank Cörvers (1999), The Impact of Human Capital on International Competitiveness and Trade Performance of Manufacturing Sectors, Maastricht: Research Centre for Education and the Labour Market (ROA).
3. Ben Kriechel (2003), Heterogeneity among Displaced Workers, Maastricht: Research Centre for Education and the Labour Market (ROA).
4. Arnaud Dupuy (2004), Assignment and Substitution in the Labour Market, Maastricht: Research Centre for Education and the Labour Market (ROA).
5. Wendy Smits (2005), The Quality of Apprenticeship Training. Conflicting Interests of Firms and Apprentices, Maastricht: Research Centre for Education and the Labour Market (ROA).
6. Judith Semeijn (2005), Academic Competences and Labour Market Entry; studies among Dutch graduates, Maastricht: Research Centre for Education and the Labour Market (ROA).
7. Jasper van Loo (2005), Training, Labor Market Outcomes and Self-Management, Maastricht: Research Centre for Education and the Labour Market (ROA).
8. Christoph Meng (2006), Discipline-Specific or Academic? Acquisition, Role and Value of Higher Education Competencies, Maastricht: Research Centre for Education and the Labour Market (ROA).
9. Andreas Ammermüller (2007), Institutional effects in the production of education: Evidence from European schooling systems, Maastricht: Research Centre for Education and the Labour Market (ROA).