2004/11 | Rapporter | Reports

Torbjørn Hægeland, Lars J. Kirkebøen, Oddbjørn Raaum and Kjell G. Salvanes

Marks across lower secondary schools in Norway

What can be explained by the composition of pupils and school resources?

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

Rapporter	I denne serien publiseres statistiske analyser, metode- og modellbeskrivelser fra de enkelte forsknings- og statistikkområder. Også resultater av ulike enkeltundersøkelser publiseres her, oftest med utfyllende kommentarer og analyser.
Reports	This series contains statistical analyses and method and model descriptions from the different research and statistics areas. Results of various single surveys are also published here, usually with supplementary comments and analyses.

ISBN 82-537-6604-1 Trykt versjon
ISBN 82-537-6606-8 Elektronisk versjon
ISSN 0806-2056

Emnegruppe
04

Design: Enzo Finger Design
Trykk: Statistisk sentralbyrå/220

Standardtegn i tabeller	Symbols in tables	Symbol
Tall kan ikke forekomme	Category not applicable	.
Oppgave mangler	Data not available	..
Oppgave mangler foreløpig	Data not yet available	...
Tall kan ikke offentliggjøres	Not for publication	:
Null	Nil	-
Mindre enn 0,5 av den brukte enheten	Less than 0.5 of unit employed	0
Mindre enn 0,05 av den brukte enheten	Less than 0.05 of unit employed	0,0
Foreløpig tall	Provisional or preliminary figure	*
Brudd i den loddrette serien	Break in the homogeneity of a vertical series	—
Brudd i den vannrette serien	Break in the homogeneity of a horizontal series	|
Desimalskilletegn	Decimal punctuation mark	,(.)

Abstract

Torbjørn Hægeland[1], Lars J. Kirkebøen[2], Oddbjørn Raaum[3] and Kjell G. Salvanes[4]

Marks across lower secondary schools in Norway
What can be explained by the composition of pupils and school resources?

Reports 2004/11 • Statistics Norway 2004

The aim of this report is provide background information for how to construct informative performance indicators for schools at the Norwegian lower secondary education level ("Ungdomsskolen") based on pupils' achievement as measured by their marks. It is commonly accepted that "school quality", however defined, may have great importance for how much pupils learn. The question is whether it is possible to quantify how schools differ in their contribution to pupils' learning. There are many other factors than the school itself, such as pupil composition and resource use, that can possibly explain differences in marks across schools. Whether and how to adjust for such factors when comparing school performance is not obvious. The relevant adjustment procedure depends on the question(s) asked and availability of data. Construction of such measures places great demands on the data, and the ideal solution may not always be feasible. This report analyses these issues in more detail, with particular focus on how reliable school performance measures may be constructed given the present availability of data in Norway. We focus our discussion around the following questions: Do differences in marks between schools reflect "real" differences or random noise? What is the impact of family background on the school results of individual pupils and differences between schools? Are differences between schools statistically significant? Do resources at schools have an impact on the performance of pupils?

Acknowledgement: This is the report from the project 'Evaluering av indikatorer for kunnskapsnivå og læringsutbytte' ('Evaluation of school performance indicators'), financed by the Ministry of Education and Research. Helpful comments from Marie Arneberg, Ådne Cappelen, Grethe Hovland, Svein Longva and Leiv Solheim as well as seminar participants at the Ministry of Education and Research are gratefully acknowledged. The authors also wish to thank Division for education statistics, Statistics Norway, for data assistance.

1 Statistics Norway, e-mail: torbjorn.haegeland@ssb.no
2 Statistics Norway
3 Ragnar Frisch Centre for Economic Research, e-mail: oddbjorn.raaum@frisch.uio.no
4 Norwegian School of Economics and Business Administration, e-mail: kjell.salvanes@nhh.no

Sammendrag

Torbjørn Hægeland[1], Lars J. Kirkebøen[2], Oddbjørn Raaum[3] and Kjell G. Salvanes[4]

Karakterforskjeller mellom ungdomsskoler i Norge:
Hva kan forklares av forskjeller i elevsammensetning og ressursbruk?

Rapporter 2004/11 • Statistisk sentralbyrå 2004

Formålet med denne rapporten er å gi bakgrunnsinformasjon om hvordan man kan konstruere informative resultatindikatorer for norske ungdomsskoler, basert på elevenes resultater målt ved karakterer. Det er allment akseptert at "skolekvalitet", hvordan man enn definerer det, kan ha stor betydning for hvor mye elevene faktisk lærer. Spørsmålet er om det er mulig å kvantifisere forskjeller mellom skoler når det gjelder deres bidrag til elevenes læring. Det er mange andre faktorer enn selve skolen, slik som elevsammensetning og ressursbruk, som også kan bidra til å forklare karakterforskjeller mellom skoler. Hvorvidt og hvordan man skal kontrollere for slike faktorer når man sammenligner skolers resultater er ikke åpenbart. Den relevante metoden for å justere skolenes resultater avhenger av i hvilken sammenheng evalueringen skal benyttes og av hva slags data som er tilgjengelige. Konstruksjon av slike resultatmål stiller store krav til datamaterialet, og den ideelle løsningen er ikke alltid mulig. Denne rapporten gir en detaljert drøfting av disse spørsmålene, med spesiell fokus på hvordan man kan lage pålitelige resultatmål for skoler gitt de data som pr. i dag er tilgjengelige i Norge. Vi fokuserer diskusjonen rundt følgende spørsmål: Reflekterer forskjeller i karakterer mellom norske ungdomsskoler reelle forskjeller eller tilfeldig støy? Hva er effekten av familiebakgrunn på enkeltelevers skoleresultater og på forskjeller mellom skoler? Er forskjeller mellom skoler statistisk signifikante? Har forskjeller i ressursbruk mellom skoler innvirkning på elevenes resultater?

Dette er en rapport fra prosjektet "Indikatorer for kunnskapsnivå og læringsutbytte", finansiert av Utdannings- og forskningsdepartementet. Forfatterne takker for kommentarer og innspill fra Marie Arneberg, Ådne Cappelen, Grethe Hovland, Svein Longva og Leiv Solheim, samt deltakere på et seminar i Utdannings- og forsknings-departemetet. Takk også til Seksjon for utdanningsstatistikk, Statistisk sentralbyrå, for data-assistanse.

1 Statistisk sentralbyrå, e-post: torbjorn.haegeland@ssb.no
2 Statistisk sentralbyrå
3 Frischsenteret, e-post: oddbjorn.raaum@frisch.uio.no
4 Norges handelshøyskole, e-post: kjell.salvanes@nhh.no

Contents

1. Introduction

Over the last decade, adoption of accountability systems for schools has become more common across countries (Kane and Staiger, 2002; Goldstein and Spiegelhalter, 1996; Hanushek and Raymond, 2004). The idea behind accountability systems is that by disclosing information about school performance and perhaps connect rewards to performance, teachers and administrators may respond by increasing their efforts to improve performance. School accountability systems usually include three elements: testing students, public reporting of school results and in some cases rewards and sanctions connected to school level performance. In addition to influencing motivation and behaviour at the school level, we would also expect that disclosure of information about school performance will rectify an important market failure – asymmetric information about school performance – and thus increase efficiency in resource use. The information about school performance is both asymmetric between individual schools on the one hand and school administrators at different levels on the other, and between schools and parents.

Whatever the ambition for an accountability system, whether it is merely to improve the information about school performance, or it is to allocate resources directly on the basis of this information, there is a need to construct reliable measures of school performance. Such indicators ought to reflect the real performance of schools and not factors more or less beyond the schools' control, such as differences in pupil composition or random noise.

The aim of this report is provide background information for how to construct informative performance indicators for schools at the Norwegian lower secondary education level (”Ungdomsskolen”) based on marks achievement of the pupils. A wide range of indicators are planned to be made publicly available via a website (“Skoleporten”) and some of the suggested indicators intend to measure the contribution of schools to pupil achievement and learning. It is commonly accepted that "school quality", however defined, may have great importance for how much pupils learn. The question is whether it is possible to quantify how schools differ in their contribution to pupils' learning. There are many other factors than the school itself, such as pupil composition and resource use like e.g. teacher intensity, that possibly explain differences in marks across schools. Whether and how to adjust for such factors when comparing school performance is not obvious. The relevant adjustment procedure depends on the question(s) asked and availability of data as well as statistical techniques. The interpretation of adjusted school performance measures is likely to be controversial. Construction of such measures also places great demands on the data, and the ideal solution may not always be feasible. This report analyses these issues in more detail, with particular focus on how reliable school performance measures may be constructed given the present availability of data in Norway. We focus our discussion around the following questions: Do differences in marks between schools reflect "real" differences or random noise? What is the impact of family background on the school results of individual pupils and differences between schools? Are differences between schools statistically significant? Do resources at schools have an impact on the performance of pupils?

Several important obstacles exist to overcome the inherit problems in constructing school level performance measures and to distinguish noise from signal of good school practices. First, it is clear from many earlier studies, internationally and in Norway, that family background is important in explaining variation in pupil achievement (Coleman et al. 1966, Hernes and Knudsen, 1976, Knudsen, 1980, Aamodt, 1982, Lie and Turmo, 2004). Hence, an unadjusted school performance measure is strongly determined by the character of the neighbourhood which constitutes the catchment area of the school. Similar families tend to cluster in local communities, and pupils typically go to their local school up to the age 16. Second, sampling variation may provide very volatile measures. The average school size per grade in Norway is low, and therefore particularly well performing or bad performing pupils may have a strong impact on a school’s performance from one year to the next. In particular, one will expect high variability or noise among small schools and they are expected to be overrepresented both among high performing schools and low performing schools. Third, in general we expect to observe substantial performance variability from one year to the next, even if we ignore the small schools, following from the evidence that differences between schools in pupil achievement only account for between 10-15 percent of the total variance in performance.[1] It is therefore important to test whether differences between schools are statistically significant. Furthermore, measures based on more than one year's performance are less exposed to volatility and they will reflect the schools' true performance to a greater extent.

The main purpose of this report is to analyse the degree to which an adjusted indicator, controlling for the family background composition of students, will provide a more reliable measure of school "quality". We also look at the persistence of school performance over time and the statistical significance of differences across schools. Finally, we study whether the adjusted school performance is related to school resources, teacher turnover and other input factors that may explain performance.

We use a Norwegian data set which includes detailed information about pupils as well as school characteristics. The data set covers the cohorts graduating from all lower secondary schools in the years 2001 to 2003 and includes information about the pupils' family background such as parental income, wealth, transfers and education, unemployment history, country of origin etc. We match this information with which school every single pupil attended and her/his marks in 11 subjects by the end of compulsory schooling (10th grade). Focus is on grade points ('grunnskolepoeng') which is the un-weighted sum of subject-specific marks, but separate analyses are also carried out for core subjects and both marks awarded for classwork and exam results. Due to the focus on academic performance, the contribution of schools in providing the pupils with non-cognitive skills and the possible trade-off is not addressed here. However, performance in non-cognitive skills such as physical education and home economics are included through marks in these subjects.

Our analysis proceeds as follows. First, we assess the effect of family background on the pupil's grade points in 10th grade (typically around age 16). We discuss the impact of school performance controlling for pupil composition and construct different indices for school performance. Secondly, we carefully assess the performance of different measures by controlling for school size, testing the significance of pair wise differences across schools and test stability in performance indicators over time. Thirdly, we add in resource use such as class size etc for each school as well as measures describing the qualifications, tenure and work experience among teachers in each school. The broader aim of the paper is then both to assess the impact of family background on student's grade points as well as the impact of school inputs on school performance.

We use a two-step procedure in our estimation strategy to construct an index for school performance at the Norwegian lower secondary education level. First we condition the distribution of 'raw' grade point averages for schools for the years 2001 through 2003 on family background variables. By doing this we also evaluate the importance of a large set of family variables on student performance. We also make comparisons and assess the stability in ranking between the raw grade point average as and the one adjusted for compositional differences. After adjusting the distribution of average grade points of schools for family composition, the second step is to asses the impact on a variety of school inputs.

The rest of the paper unfolds as follows. In chapter 2 we provide some background and a brief discussion of methodological issues. Chapter 3 describes the data sources. In Chapter 4, we outline our empirical specification. Chapter 5 analyses how individual school achievement varies with family background. Chapter 6 to 8 go in detail on indicators of school performance: The importance of adjusting for differences in family background, the stability of measures over time and the statistical significance of differences between schools. Chapter 9 analyses how measures of school performance vary with school characteristics like teaching hours and teacher composition. Our conclusions are given in Chapter 10.

[1] In fact differences across schools in Norway is at the lower end compared to other countries in that it can only explain 10 percent of the variance in school performance (in reading) using data from the PISA survey (Lie and Turmo, 2004).

2. Background and previous literature

The three main research issues in this study is how to compare performance of schools by constructing adjusted measures of school performance, how to assess reliability and persistence of school performance measures and finally how to relate school performance to resources used by the school. Different strands of the education literature are relevant here. In this chapter, we briefly discuss three of them. First, there exists a large amount of evidence on how family background is an important determinant of student achievement. Thus, since student composition with respect to family background differs across the schools' catchment areas, it is of crucial importance to adjust for students' family background when constructing school performance measures used to compare schools. Second, a small and recent literature deals with other aspects of comparing school performance. This literature points to the volatility of different school rankings, and how the ranking of schools can be influenced by other factors than student composition and those which possibly can be affected by the school (management or teachers). Third, there is a relatively large literature dealing with school efficiency and how school resources are related to student achievement.

2.1. The importance of controlling for students' family background

The modern literature on the effects of school resources and organization, the importance of student composition and family background on student achievement, started with the so-called Coleman Report in 1966 (Coleman et al., 1966). Coleman and his group of researchers collected nationwide data for the US on school inputs, socio-economic background of students and test scores. The study was mandated by the 1964 Civil Rights Act, and had a strong impact of the desegregation of the US schools. The main lesson from this study was: 'It's all in the family'. School inputs had very little measurable impact on student achievement, but the effect of family background and the composition of the student body had dominating effects. In fact, school differences in student performance only accounted for 10-15 percent of the total variance. Thus, the study showed that when comparing schools with respect to performance, it is of vital importance to adjust for family background which could explain far more of the variance in student achievement.

Following Coleman et al.'s contribution, a large literature was spurred analysing the impact of family background on school performance. For instance, in Norway Hernes and Knudsen (1976) provided results similar to those found in the original Coleman report, both on family background and school resources. There is consensus in the international literature that family background is important in explaining student performance both in terms of test scores, adult educational attainment and adult wages.[2] However, it varies across countries which family background variables are important. For instance in some countries parents' economic background is found to be important while parents' educational background or family employment history are found to be of greater importance in other countries. In Norway, previous studies suggest that parents' educational background is important, while economic resources explain very little of adult educational attainment, see e.g. Raaum, 2003. When it comes to other family background variables, such as parents' and particularly the mother's employment history, and their importance for children's social and cognitive development, the results varies from mother's employment having a negative to a positive effect (see e.g. Hanushek, 1996; and Baydar and Brooks-Gunn,1991, Shonkoff and Phillips, 2000).

However, the interpretation of the impact of family background, and thus how to draw policy implications, is not obvious. Although in the present context we are not primarily concerned with different explanations of *why* family factors are important, we provide a brief overview of the main questions discussed in the literature. One important issue is causality. In the case of parental education, why do children of well

[2] See Haveman and Wolfe, 1995 for an international overview; and Raaum, 2003 for an overview of the most recent Norwegian results on educational attainment. OECD (2000) and Lie and Turmo (2004) discuss recent result for Norway using survey data form the PISA study: See Woessman (2004) for a comparative study of the TIMSS data including Norway and Bakken (2003) for a study of pupil's achievement with focus on ethnic minorities.

educated parents perform better in school? A causal effect means that an increase in parents' education level improves the performance of children, implying that policies that raise educational attainment of one generation spill over to the next. However, the correlation between parents' education and children's test scores may reflect that clever parents get clever children who do well at school. Or alternatively, there are other factors than education itself which creates the superior environment for children in families with highly educated adults. Causal effects are hard to identify, also in the case of family effects. Hence, it is difficult to assess the effects of policies directed at increasing the equality of educational opportunities.[3]

A second important issue is that a child's family background probably interacts with the background of other families constituting the neighbourhood where the children grow up. Other institutions in the same neighbourhood, such as preschools, may have lasting effects on child development. Hence, all of these factors must ideally be analysed together in order to identify the effect of family background. It is also clear that there is a tendency that families sort into neighbourhoods where other families of similar social background lives. So both families and neighbourhoods (both as peer influence and institutions), in addition to schools, probably also interact strongly with school inputs in explaining individual student performance. However, this potential interaction makes it very difficult to disentangle the three effects, partly because of family sorting.[4]

2.2. Constructing reliable measures of school performance

The main aim of this report is to construct a reliable measure of school performance as a part of the school accountability system in Norway. Several important problems arise which are discussed in the literature. They can mainly be categorized into two types. One problem is how to construct a school performance measure which only reflects the school behaviour and thus control for irrelevant factors and noise. The other – and related – problem is to construct measures that are not volatile and changing from one year to the next and actually have enough information to distinguish schools from each other.

Unadjusted school performance measures are strongly determined by the character of the neighbourhood which constitutes the catchment area of the school. Families tend to cluster in local communities and pupils typically go to their local school up to 16 years of age. One way to address this problem is our preferred approach where we adjust for the pupils' observed family background. An alternative way, widely recommended in the literature, is to focus on so-called value-added measures. These are based on the gain in performance for the same students from one period to the next and thereby implicitly controlling for other factors such as family background. While such measures are considered to be very favourable, they also have important drawbacks. First, students differ not only in their baseline performance, but also in their performance trajectory. It has been found that students with educated parents not only performed best the first year measured, but also improved most from one year to the next (Kane and Staiger, 2001). A more sophisticated discussion of using value added measures can be found in Todd and Wolpin (2003). In their review of the literature they note that using value added measures is a way to proxy for missing variables when the pupils' background histories are unknown. They then proceed to discuss under which conditions such proxy variables are valid. Their main suggestion is that when one is interested in analysing cognitive achievement of children, one should ideally use data for all past and present family and school inputs. In our case we would like to have data for inputs in the primary school the pupil went to as well as the students' performance there, and information on parents' resources – both monetary, educational and heritable endowments – both pre-school, at the primary school and at secondary level. This is an ideal which almost never is feasible. In practice, it is particularly hard to track the school environment of every single pupil during a ten year period or more. Todd and Wolpin (2003) discuss in detail the implication of using approximations to the ideal model.

Another problem using value added measures noted by Kane and Staiger (2001) as an alternative to control for family background (and possibility performance in earlier stages of schooling), is that value added estimates are typically imprecise and more volatile than measures based on performance levels. This is exactly the second main problem in establishing performance indicators; to establish performance indicators that are precise and do not fluctuate too much. One problem here is that sampling variation may provide very volatile measures. The average school size per grade in Norway is low. Therefore, particularly high or low achieving pupils may have a strong impact on a school's performance from one year to the next. In particular, one will expect high variability or noise among small schools; they are expected to be overrepresented both among high performing schools and low performing schools. In addition, one should in general expect great variability

[3] See Altonji and Dunn, 1996; Black, Devereux and Salvanes, 2003, for different approaches to establish causal effects of parents' education on children's' educational performance, and Cameron and Heckman, 1998; and Carneiro and Heckman, 2002 for heterogeneity of effects.

[4] See Solon (2002) for an instructive discussion of the endogeneity problems of identifying family and neighborhood effects; Raaum, Salvanes and Sørensen, 2003 a,b for analyses of family and neighborhood effects on adult schooling and earnings based on Norwegian data.

in performance over time also when small schools are ignored, since school differences in student performance only account for between 10-15 percent of the total variance in performance.[5] This makes it important to test for significant differences across schools when comparisons are made. Furthermore, one should construct measures based on more than one year's performance in order to avoid very volatile measures.

2.3. Student performance and school resources

There exists a strong agreement on the importance of family background – although the interpretation of this effect is far from straightforward. Much less agreement exists on Coleman et al.'s (1966) results regarding the impact of school resources on student performance. There seems to be agreement when it comes to the effect of some inputs and some outputs, and likewise disagreement on others. There appears to be agreement that teacher quality matters, but there is much less agreement when it comes to use of other school resources. The dispute is particularly strong regarding the impact of school resources measured by e.g. expenditures and teacher-pupil ratios on student achievement such as marks and test scores. In several studies Hanushek with co-authors do not find support for any impact of school resources on student test scores (for overviews see Hanushek, 1996, 2003). Similar results have been found for Norway measuring the effect of class size on test scores, although a weak impact of class size is found for Norway (see Bonesrønning, 2003). Krueger with coauthors have in several cases found positive effect of school resources on student test scores (Card and Krueger, 1992; Krueger and Lindahl, 2002 for Sweden). In summing up the literature by undertaking a so-called meta analysis, Hanushek (2003) concludes that the overall pattern is that no school inputs can be associated with student achievements. Krueger (2003), on the other hand argues that the meta study is biased towards the conclusion of no impact of school resources. However, the same types of school resources seem to have a stronger impact on post-school outcomes like final educational attainment and earnings (Betts, 1996, Dearden, Ferri and Meghir, 2002, Dustmann, Rajan and Van Soest, 2003, Wilson (2002)).

Similar methodological questions are relevant for assessment of the effect of school resources on student performance, as well as for the impact of family background on student performance. The main problem may be illustrated using an example. We have two schools – one using a high level of resources per student (e.g. small classes) and one with a low level of resources. Pick two identical students in the population and put one in each of the schools and test performance after a year. The difference in the performance by the pupils would then give us the answer to whether more resources have an impact on student performance. Repeating this exercise with a sufficiently large number of pupils, a causal link between school resources and individual achievement could be tested.

Since we cannot perform this type of experiments, we have to rely on other ways of solving the inherit endogeneity problem(s). First, one issue is related to how resources are allocated across schools. These processes are crucial as we have to rely on the actual, implemented, non-experimental, distribution of (teaching) resources to identify the parameters of interest. There are numerous administrative regulations and specific measures within the school sector which link the allocation of resources to a particular school and the ability or learning 'capacity' of the pupils enrolled. In short, compensatory resource allocation implies that estimates of the effects of teacher input on school performance are likely to be biased downwards. For instance, if schools respond to enrolment of weaker pupils by providing more teaching hours and smaller classes for this group of pupils, we will underestimate the effect of resource use in schools. Second, parents, teachers and pupils can adjust their behaviour in response to changes in resource use. If we concentrate on class size as a composite measure of school resources, student performance may vary across schools with small and large class sizes because of teacher adjustment and/or parents' adjustment. Teachers use small classes for less able students, parents choose neighbourhood based on school quality (class size), and schools with small class size may also have other favourable characteristics (attracting better teachers etc). See Bonesrønning (2004a, 2004b) for this type of adjustment to school resources. He finds that parents' effort towards children declines when class size is reduced, obscuring the relationship between class size and children's performance.

The main endogeneity problem has two possible solutions: 1) Use of natural experiments or instrumental variable estimators 2) Use matching estimators. There is a large literature using natural experiments or instruments. The idea is as follows. If there exists a maximum class size rule (commonly referred to as the Maimonides' rule), it will cause exogenous class size variation. Studies that use this strategy include Angrist and Lavy (1999), Hanushek, Kain and Rivkin (1998), Hoxby (2000), Browning and Heinesen (2003), and for Norway see Bonesrønning (2003). These studies usually find very small or no effects of class size. Matching is another approach to use control for endogeneity which basically means that you construct control groups and treatment groups which are similar on many observable variables (see Heckman, Ichimura

[5] In fact differences across schools in Norway is at the lower end compared to other countries in that it can only explain 10 percent of the variance in school performance (in reading) using data from the PISA survey (Lie and Turmo, 2004).

and Todd, 1998). An example of this strategy is used by Machin and McNally (2003) to find the effect of increased school resources and organization of teaching on literacy in UK schools in 1997 and 1998. They do find non-neglible effects for weak students of increased teaching resources and improved teaching practices. The effects are stronger for boys than for girls and the effects appear to be relatively persistent over time.

3. Data sources, sample and variable construction

This section describes the data sources used in this study, the construction of important variables and sample restrictions.

Our sample covers all students who completed compulsory education in Norway (10th grade in the lower secondary school) in the years 2001 to 2003. We use three main types of data in our analysis containing information about pupil marks by subject, individual characteristics and family background, as well as schools. Most of the data are taken from various administrative registers. The use of common identifiers of individuals and schools across registers facilitates matching of different registers and enables us to construct a dataset that is very rich in both individual and school characteristics.

3.1. Marks and school identification

Information on individual marks was collected by Statistics Norway for 2001 and by the Norwegian Board of Education for 2002 and 2003. The data contain a personal identification number, an identification number of the school from which the pupil graduated and information on marks. For the 2002- and 2003-cohorts we have information of marks in each subject, both marks awarded for classwork (standpunkt) and examinations. For 2001 we only have information on the total grade points (GP; grunnskolepoeng), which is an aggregate of marks in individual subjects, see below.

Pupils are awarded marks in 11 subjects: Norwegian (primary form, written), Norwegian (secondary form, written), Norwegian (oral), mathematics, English (written), English (oral), social studies, science and the environment, Christian knowledge and religious and ethical education, arts and crafts, music, home economics and physical education. In all these subjects, marks are awarded for classwork. In addition, all pupils are tested by a written examination in Norwegian, English or mathematics, and by oral examination in one of the subjects. Marks are awarded on a scale from 1 to 6, where mark 6 indicates that the pupil holds exceptionally high competence, and 1 indicates that the pupil has attained little competence in the subject. Table 3.1 shows the distribution of marks for a typical subject in 2002-2003.

While the mark for 'Norwegian', both primary and secondary form, measures the Norwegian in 'bokmål' and 'nynorsk', respectively, according to the standard curriculum for the large majority of pupils, it is actually an aggregated measure based on other subjects for a group of minority language students. In particular, according to Læringssenteret (2004), approximately 2,600 pupils have Norwegian as their second language ('Norsk som 2.språk'). The mark for this subject is included as Norwegian, primary form, and we have no information to single out the students with this alternative curriculum. The competence which is needed to obtain a given mark in 'Norwegian 2' is substantially lower than for the standard curriculum. Moreover, the mark for Norwegian, secondary form, covers the mark in their own/parents' native language for a large number of ethnic minority pupils. In short, the marks in 'Norwegian' exaggerate the average Norwegian language competence attained by pupils with immigrant background. With the current 'aggregated' data, we are not able to evaluate the importance of this.

Based on the marks awarded in individual subjects, a summary measure of total grade points ("Grunnskole-poeng" in Norwegian, henceforth GP) is constructed. GP is the basis for ranking applicants to upper secondary schools, and is calculated as follows: In each of the eleven subjects, one takes the average of marks awarded for classwork and marks for exams, and then sum over all the eleven subjects. Maximum GP is then 66. If a pupil does not have marks in all subjects (for whatever reason), marks in up to two subjects are imputed based on the pupil's average marks in the other subjects. E.g. if a pupil has marks in eight subjects only, with an average mark of 2, 4 points are added to GP, so that GP is 20 instead of 16. In our analysis based on GP, we include students with marks in at least five subjects.

Table 3.1. Marks awarded for classwork, Norwegian (primary form), written, 2003

Mark	Number of pupils	Percent
1	214	0.41
2	4,526	8.58
3	14,757	27.98
4	19,110	36.23
5	12,767	24.21
6	1,371	2.60

Our focus is on academic achievement and the contribution of schools in providing pupils with non-cognitive skills is not addressed here. Given the emphasis on academic achievement and hence performance measures based on marks, it is not obvious how individual marks should be aggregated into one summary measure. It may be argued that some subjects provide more information about overall competence than others, and that marks in these subjects are better predictors of later success in the educational system and in the labour market. In this respect, a summary measure giving equal weight to e.g. mathematics and home economics may not be the most adequate. However, apart from experimenting with some alternative weighting schemes later in the report, we do not pursue this topic further, but use GP as our main measure of individual (and school) performance. GP forms the basis for admissions into upper secondary schools and has in some regions like Oslo been the sole criterion for some years. Hence, GP is a relevant summary measure as it implicitly reflects the priorities of school administration and politicians when it comes to the relative importance of different subjects.

3.2. Pupil characteristics and family background

By combining a large number of administrative data sources, we have assembled detailed information on pupil characteristics and family background for all pupils who completed compulsory education (10th grade in the lower secondary school) in the years we study. This provides us with a wide range of family background characteristics. In addition to basic demographic information, we have information of parents' education, immigrant status, parents' wealth, income and (un)employment histories, disability status and receipt of social assistance. Below, we describe the variables in more detail.

Demographic information

We construct dummy variables for the pupil's gender, quarter of birth (given graduation in the year they turn 16) and graduation in years earlier than expected.

Family structure

The following sets of variables are included

- Parents' marital status - dummy variables reflecting whether they are married (to each other), cohabitants, separated, divorced or neither of these.
- Unknown parents - dummy variables indicating whether the father and/or mother is unknown
- The age of the mother and father at the birth of their first child - dummy variables reflecting age intervals
- The number of full siblings and the pupil's rank in the birth order - detailed set of dummies
- Half siblings - dummies indicating the number of half siblings

Parents' education

We have information of the highest completed level of education for each parent. Parents' education is classified into five categories: Lower secondary (up to 9 years of schooling), upper secondary (10-12 years), lower tertiary (13-16 years) and higher tertiary education (17 years or more). We also include an additional category for missing education information. Since missing education information typically appears for immigrants, missing information is also interacted with immigrant status. Based on this classification, we construct dummy variables for all combinations of father's and mother's education.

Immigrant status

Pupils who were born abroad by non-Norwegian parents and pupils whose parents were born abroad are classified as immigrants in our analysis. We construct a detailed set of dummy variables indicating country/region of origin (Scandinavia, ex-Yugoslavia, other Eastern Europe, Turkey, Western Europe (plus USA, Canada, New Zealand, and Australia), Somalia, other Africa, Sri Lanka, Iran, Iraq, Vietnam, Pakistan, Vietnam, other Asia (plus Oceania), Latin America). We also control for the age of immigration *for the pupil*, with dummy variables distinguishing between those who born in Norway or immigrated before they were three years old, and those who immigrated when they were 3 to 5, 5 to 7, 7 to 9, 9 to 11, 11 to 13 or 13 years or more. We experimented with more detailed country classifications, and also with distinguishing between first- and second-generation immigrants on country/regional level, but the results were basically unaffected.[6]

Economic resources

Based on information about individual taxable labour income ('pensjonsgivende inntekt'), we calculate a measure of family income for the pupil as the sum of the father's and the mother's income during the last ten years (regardless of marital status). Although income tends to be quite persistent over time, we have included not only current but also previous income, to make it reflect family income over the period the pupil went to school, and not only around graduation. We then construct dummy variables reflecting the position

6 Interpretation of the difference between first and second generation immigrants is hard when country/region of origin composition differ across groups like in Arnesen (2003).

(quintile) in the family income distribution (for the parents of graduating pupils, not the population as a whole).

Based on information of individual taxable wealth, we calculate a measure of family wealth for the pupil as the sum of the father's and the mother's wealth for the year prior to graduation (regardless of marital status). Wealth typically increases over the major part of the life cycle. We therefore construct age-specific wealth distributions, where we divide families into five-year age intervals based on the average age of the parents. We then construct dummy variables indicating whether the family belongs to one of the upper four deciles of its respective age-specific wealth distribution. Note that a majority of families have negative taxable wealth, since their net financial wealth is negative. The tax value of housing, which is most common non-financial asset, is far below market value. Negative taxable wealth is reported as zero.

Unemployment, disability pension and social assistance

We use information of the parents' unemployment history to construct variables indicating the incidence of unemployment during the ten years prior to the pupil's graduation. We define a person to be unemployed in a given year if he or she was registered as unemployed for at least three months of a calendar year. Based on this definition, we construct dummy variables, separately for mother and father, for unemployment in the graduation year, and for unemployment one, two, three, four and five or more years during the ten-year period prior to graduation.

Similarly, we construct variables indicating the receipt of disability pension and social assistance. We define a person to be on disability pension in a given year if he or she received disability pensions for more than six months of the calendar year. Our criterion for defining a person as receiving social assistance is that he or she received at least Nkr 20,000 (approx € 2,500) in a given year. The dummy variables for disability pensions and social assistance are constructed in the same manner as for unemployment.

3.3. School characteristics

Our two main sources of information on school characteristics are the Compulsory School Information System ("Grunnskolens informasjonssystem" in Norwegian, henceforth GSI), and individual information underlying Statistics Norway's teacher statistics.

From GSI, we utilize information of the number of pupils in grade 8, 9 and 10, the number of classes by grade and the total hours of instruction for pupils at grade 8-10. From this information we construct different measures of resource use, such as hours of instruction per pupil, which may be decomposed into the product of teacher hours per class and classes per pupil. The resource variables are calculated as the average of the three years prior to graduation (8-10th grade) to reflect the resource use in the whole period the pupil attended lower secondary school, and not only the final year. (More details on the construction of resource variables are given in Chapter 9). In addition to resource use variables, GSI also provides us with information of whether the school is privately owned and whether the school is a combined primary and lower secondary school containing pupils from 1-10th grade as opposed 8-10th grade lower secondary schools.

From the teacher statistics, we use detailed individual information on teachers to construct measures reflecting level of formal qualifications, age profile, gender composition and turnover of the teacher staff. Unfortunately, teacher characteristics are not matched to what grade or class the teacher worked with. Again, these variables are calculated as averages over the three years prior to graduation. Private schools are not included in the teacher statistics.

3.4. Sample trimming

Table 3.2 shows details of the sample trimming procedure for the graduation year 2003. It provides information on sample sizes at different levels and where observations are lost due to missing data. We have information on marks for more than 55,000 individuals. Around 4.5 percent are excluded form the estimating sample. Missing information of family background is the major reason for excluding individuals from the sample. But also a significant number of pupils with too few marks and without information on which school they attended are excluded from the sample. Our final sample in individual regressions consists of nearly 53,000 pupils from almost 1,100 schools. In school level analyses, missing information at the school level removes up to nine percent of the schools.

Table 3.2a. Sample trimming for individual level regressions, 2003

Total number of observations	699,669 single-grade-observations, from 55,231 different individuals
- invalid personal identification numbers	838 single-grade-observations, removing 93 individuals
= total number of individuals with marks / grade points	55,138
- no information on family background	1,661
- missing information on school attended	229
- too few marks (less than 5 subjects)	483
- no marks awarded for classwork	10
= sample for grade point regressions	52,755 individuals from 1,094 different schools
of which are going to medium-sized or large schools (>30 pupils in 10. grade)	44,329 individuals from 551 different schools
of which are going to large schools (>60 pupils in 10. grade)	32,499 individuals from 322 different schools

Table 3.2b. Sample trimming for school level regressions, 2003

All schools in grade point regressions	1,094 different schools
- lacking information on resources, type of school or number of pupils	45 schools (1,049 remaining)
- lacking information on teacher characteristics	41 schools (1,008 remaining)

Table 3.2c. Sample sizes for individual level regressions on single subjects. Sample trimming procedure similar to Table 3.2.a

Norwegian (primary form, written), total	52,658
Norwegian (primary form, written), classwork	52,521
Norwegian (primary form, written), exam	11,726
Mathematics, total	52,504
Mathematics, classwork	52,459
Mathematics, written exam	20,116
English, total	52,332
English, classwork	51,798
English, written exam	19,307
Social studies, total	52,660
Social studies, classwork	52,632
Social studies, oral exam	7,356
Physical education, classwork	52,270

4. Estimation strategy

Our estimation strategy contains three steps. First, we estimate the impact of gender and family background variables on individual achievement. The main focus is on grade points, but we also look at five specific subjects. Second, we construct and evaluate school performance measures, based on the average grade points (GP) at the school level. We compare the (unadjusted) school GP average with a measure adjusted for pupil composition using a large number of family characteristics. Technically, the adjusted measure of school performance is a school fixed effect from a regression of individual GP on a large set family characteristics and school dummies. The stability over time in the ranking of schools and the effects of pupil composition adjustment is assessed, along with thorough testing of whether schools are significantly different. With data for several cohorts, we are also able to discuss the persistence of between-school-differentials. If school quality really explains the ranking of schools by average GPs, we would expect schools not to move very far from one year to the next. Finally, the third element in our procedure is to evaluate the impact of a variety of school inputs on school performance measured by the family background adjusted school mean GP. This section describes the empirical models and some technicalities, while the results are reported in Chapter 5 to 9.

4.1. Grade points and family background

We start out by studying the effect of family background variables on individual pupil grade points (GP), by estimating the following simple equation without conditioning on school effects[7]

(1) $GP_i = \beta F_i + \varepsilon_i$

where F_i is a vector of family background variables (as described in Section 3) for pupil i consisting of parents' education, income, wealth, country of origin, marital status, incidence and history of unemployment, social assistance and disability pensions,, the number of siblings, parents' age at first childbirth, the pupil's gender etc. The β-vector is the estimated effects of the family background variables. ε_i measures the unobserved individual characteristics as well as school effects and random components. The equation is estimated without controlling for regional fixed effects. There may be an argument for including possible regional effects, capturing e.g. difference in marking practice as since it is sometimes claimed that different practices in setting marks exist when regional examination commissions are used. However, it is unlikely to make a big difference as is shown in Arnesen (2003). She finds that mark differentials in core subjects across counties are small, conditional on parents' education.

By undertaking a simple ANOVA analysis of the family background variables by including sets of individual background variables sequentially, we can obtain measures of how much the different components of the family background of students explain the variation in grade points across pupils.

In addition to assess the impact of family background on grade points for each student we also assess an alternative achievement measure. Rather than grade points which are un-weighted means of all eleven subjects, we apply the schools' own weighting of the subjects by using the hours of instruction in each subject for all the three years spent in lower secondary school.[8] A nationwide plan for time allocation for each field is provided by the Ministry of education and research (Lærerplanen 1997). By using this weighting scheme core subjects such as mathematics are given a higher weight than for instance home economics. Results from examinations and marks awarded for classwork are still given equal weight.

[7] We actually estimate this equation also by including school fixed effects and very similar results are obtained. The model without school fixed effects is used for convenience as the CPU-time increases dramatically when school dummies are included. To avoid programming and to get results within reasonable time horizon, we opted for the simple model.

[8] We actually also use hours of instruction in the eleven subjects in both primary and lower secondary school (1-10th grade) as weights, but the differences to the scheme based on 8-10th are negligible.

We also evaluate the impact of parents' background on results in individual subjects such as mathematics, English and Norwegian as well as physical education and social studies. We expect the impact of family background to be stronger in the three core subjects and less so for physical education. Furthermore, we assess separately exam results and average marks from tests and activity in class over the whole year. It is not clear which is preferable in measuring pupils' knowledge. Obviously, examination marks may provide a good signal of the pupils' knowledge, since likes and dislikes of the teacher do not affect the outcome. On the other hand, the result depends only on one test and the student may have had a bad or good day. The results based on performance over a whole year may be a better predictor for the students' knowledge.

4.2. School performance controlling for student composition

In order to obtain an estimate of school performance, conditional on the family background composition of the pupils, the following equation is estimated:

$$(2) \quad GP_{ij} = \beta F_i + \Sigma q_j S_{ij} + \varepsilon_i$$

Where GP_{ij} is grade points for pupil i in school j $(j=1,\ldots,J)$, F_i is the vector of family background variables described in the previous section. (F_i is measured as deviations from sample means) S_{ij} is and indicator variable which is one if pupil i is in school j and zero otherwise, and q_j is the school fixed effect. ε_i is the error term. The estimated school fixed effect ($\hat{q}_j$) can be interpreted as the average GP, conditional on family characteristics being equal to the sample mean. These school fixed effects can be compared with the unconditioned school effects, GP_j, which is computed as the average of the grade points within in each school:

$$(3) \quad GP_j = \frac{1}{n_j}\sum_{i=1}^{n_j} GP_{ij} \text{ , } j=1,\ldots..,J$$

By comparing the unadjusted school means (GP_j) and the estimated adjusted school fixed effects, ($\hat{q}_j$), we can assess the extent to which sorting on family background affects the ranking of schools.

As we discussed in chapter two it is of vital importance not only to compare the stability of adjusted and unadjusted school performance measures, but also to be careful in testing for statistical differences by undertaking pair wise statistical significance tests for all schools. In addition to test the stability across adjusted and unadjusted performance measures, and across years, we analyse in detail whether differences in grade points across schools can be considered as pure noise or important and significant differences. We also take into account that small schools are much more volatile in performance because small changes in the composition of children may make a big difference in the average performance of each school. This is the reason why small schools tend to appear in the tails of the school performance distribution.

A short comment on our specification is due. A preferred approach to control for the background of pupils and therefore to isolate the effect of schools in providing pupils with cognitive skills and therefore distinguish between schools, will include family background at different stages – preschool and probably pre-birth resources of parents -for the pupils as well as for the impact of earlier school experience – including preschool - of the pupils (Todd and Wolpin, 2003). Our specification for the most part ignores the history and timing for resources at different levels in the children's development. Hence, our approach as well as the value added approach has shortcomings and will potentially provide biased results. However, our rich set of family background variables is expected to be very helpful in controlling for most previous experience, as family background variables tend to be highly correlated over time.

4.3. Do school resources matter?

Equation (2) decomposes the individual performance measure as a result of family background factors and schools. The school effects capture observed as well as unobserved characteristics of the school. Typical observables include teachers' education, class sizes, administrative resources etc. Examples of (for us) unobserved characteristics are efficiency in organizing the use of a given set of resources, and also the different types of peer effects between teachers and pupils and between pupils. In this second step we retain the estimated school fixed effect, ($\hat{q}_j$), from section 4.2 and test whether these school effects can be explained by different observed school inputs in the following model:

$$(4) \quad \hat{q}_j = \gamma K_j + u_j$$

where K_j is a vector of observed school characteristics. We focus in this analysis on two main categories of school inputs: 1) School resources (including class size and teacher density), 2) Characteristics of the composition of teachers (formal education of teachers, experience, tenure and turnover).

Hence, by estimating the model given in equation (4) we obtain the effect of the different school inputs on the adjusted school effects, i.e. when pupils' (observed) family background are controlled for. However, our specification and procedure for estimating the effect of family background variables and in equation (1), and the estimation of the product function at the school level in equation (3), require some comments. The error term, u_j in equation (4), captures the unobserved characteristics of schools as well the impact of the average unobserved pupil characteristics. This means that unobserved pupil characteristics like previous school history as well as unobserved family characteristics are correlated with our error term, may bias our results. But note that we use an unusually rich set of family background variables for each student to control for pupil characteristics. However, the fact the error term in equation (4) is correlated with unobserved school characteristics will bias our results and we cannot under general conditions estimate causal effects of school resources on student performance. For example, resources to schools may be given to compensate for pupils with special needs who require extra follow up. This means that small classes and more intensive use of teachers tend to take place in schools where pupil achievement is low. In other words, since we do not have all the information about which schools have more demanding pupils and how the compensating resource allocation operates, we will tend to underestimate the effect of resource use on student performance. Some of this information is probably captured by the rich set of family background variables and demanding pupils. However, we probably do not capture all. All in all we expect our resource coefficients to be downward biased.

5. Pupil achievement, gender and family background

In this chapter, we describe how individual achievement varies across pupils with different family background and gender. Achievement is measured by the grade points (GPs) at the completion of lower secondary school as explained in Chapter 3. First, we present the unadjusted distributions of GPs for girls and boys, as well as by family background characteristics, using data for the 2002- and 2003 cohorts. Second, we discuss and provide evidence on the marginal contributions of each set of family variables in explaining the overall variation in grade points. A standard ANOVA decomposition is used for this purpose. Third, we calculate the predicted grade points for children growing up in families with specific sets of family characteristics. A comparison of these archetypes provides information on the marginal 'effects' of certain sets of family characteristics. We also consider the marginal 'effects' of a number of single variables like birth-order, age at enrolment in school etc. Fourth, in addition to using grade points, which is an un-weighted sum of all eleven subjects, averaged across exam marks and marks awarded for class-work ('standpunkt-karakter')[9], we asses the impact of family factors and gender on selected subjects. We test the hypothesis that learning in the family during early childhood and adolescence as well as own and parents' cognitive endowments have a larger impact on core subjects like mathematics and Norwegian than in subjects such as physical education. Finally, an alternative aggregate measure of individual achievement can be based on the time allocated to different subjects. Teaching hours can be interpreted as implicit weights attached to each subject, i.e. the school system's own weighting scheme. Using these alternative weights, we test whether the impact of family background depends on how we aggregate performance across subjects into a one-dimensional measure.

The family matters. We show that performance in school, measured by e.g. GP, varies substantially between pupils from different family environments.

[9] At the end of the school year, marks for classwork (including tests) are awarded for subjects completed that year, reflecting the degree of competence attained in accordance with the curriculum for the subject concerned.

The reproduction of educational attainment has been subject to intense research for decades in the social sciences, see Aamodt (1982) for an early study of Norwegian data. Large, positive associations between parental educational attainment and children's achievement/ behaviour are among the most substantial and replicated results in studies of child development, starting with the Coleman report from the late 1960s; see e.g. Shonkoff and Phillips (2000) for a recent survey.

For the purpose of this study, we *do not discuss why* the various family characteristics matter. Since our main focus is on why *schools* perform differently, an adjustment for pupil composition is needed. Whether well-educated parents provide a superior learning environment at home, express effective expectations or simply transmit genes or cultural values that promote school performance is not addressed in this study. In other words, the distinction between casual effects or confounding factors is not discussed here; see e.g. Black, Devereux and Salvanes (2003) for a test of parental education as a causal effect on their children's educational performance using Norwegian data.

5.1. Grade points across groups

On average, girls outperform boys, see Table 5.1. The mean differential is 4.6 GPs in the favour of girls. In other words, girls have on average 0.42 (=4.6/11) higher marks in each of the eleven subjects, compared to boys. The percentiles indicate that the gender differential is large throughout the GP distribution and somewhat larger among low-scoring pupils. However, the higher girl average is far from fully explained by a large number of boys who score particularly low. The fact that girls perform better is well known from previous Norwegian studies of survey data. For instance Hernes and Knudsen (1976), using data from the early 1970s, found a large gender differential in exam result at the lower secondary education level. There is also abundant international evidence showing that girls outperform boys in the class room. Our results for Norway is in line with this pattern, see Machin and McNally (2003) for evidence on this and tests of different explanations.

Table 5.1. Grade point means. By gender, family characteristics and immigrant background

Type of pupil	Mean	# obs	Percentiles				
			90 %	75 %	50%	25 %	10 %
All	43.31	106.494	54.67	50.50	44.17	37.00	30.50
Gender							
Girl	45.64	52.129	55.83	52.33	46.93	39.97	33.50
Boy	41.08	54.365	52.83	48.00	41.50	34.67	28.67
Family characteristics							
Low educated parents	36.70	2.656	48.50	43.17	36.48	30.33	25.33
Medium educated parents	41.91	36.027	53.00	48.50	42.33	36.00	30.25
Highly educated parents	51.38	2.033	59.17	56.00	52.50	47.75	42.50
'Poor' parents	38.08	13.607	50.67	45.00	38.00	31.50	26.00
Rich parents	48.35	9.555	57.00	54.00	49.50	43.67	37.83
Married parents	44.91	65.703	55.33	51.67	46.00	39.00	33.00
Divorced parents	40.78	18.207	53.00	48.00	41.00	34.10	28.33
Separated parents	41.86	3.248	53.67	49.33	42.50	35.25	29.00
Unemployed parents	39.75	5.806	52.17	47.17	40.00	33.00	27.00
Parents without unempl. history	44.62	72.720	55.33	51.50	45.65	38.67	32.27
Parent(s) have unempl. history	40.49	33.774	52.50	47.50	40.88	34.00	28.17
Parent(s) with disability pension	40.23	11.177	52.67	47.50	40.50	33.50	27.50
Parents without disability history	43.70	94.652	54.83	50.83	44.50	37.50	31.00
Parent(s) have disability history	40.21	11.842	52.67	47.50	40.50	33.37	27.50
Parent(s) recent social assistance	38.16	9.883	51.33	45.65	38.00	31.00	25.50
Parents without social ass. history	44.65	82.659	55.00	51.50	45.50	38.75	32.63
Parent(s) with social ass. history	38.69	23.835	51.50	46.00	38.50	31.83	26.22
Immigrant background							
Immigrant (non-western) girl	40.42	2.728	53.00	47.85	41.07	33.61	27.13
Immigrant (non-western) boy	36.73	2.906	49.50	44.00	37.00	29.83	23.65
Norwegian background girl	45.92	48.799	56.00	52.50	47.17	40.25	34.00
Norwegian background boy	41.33	50.894	53.00	48.25	41.83	35.00	29.00

Parental education is the strongest predictor of pupil performance. The GP difference between the two extremes, both parents with high university degree (17 years of schooling or more) versus both with compulsory education only, is close to 15 GPs. Children of parents with upper secondary education ('videregående skole', 10-12 years) score, on average, 5 GPs more than the lowest parental education group.

Achievement is highly correlated with economic resources of the family. The GP differential between pupils from 'poor' and rich families is about 10 GPs. 'Poor' parents means that they are located in the lowest quintile (20%) in the parental earnings distribution and do not have financial wealth among the wealthiest 30%. Rich parents are in the highest earnings quintile and among the ten per cent with highest financial wealth.

United parents have children who perform better in school. Pupils of divorced parents score, on average, 4 GPs less than those with parents married to each other and about 1 GP less than pupils with legally separated (not yet divorced) parents.

Parental unemployment, social assistance and, to some extent, disability pension among the parents are all family characteristics that are negatively correlated with pupil performance. For example, children of parents who experienced unemployment (defined as duration of three months or more per year) during the last ten years achieved 4 GPs less than those of parents without any unemployment experience.

Immigrants[10] from non-western countries score, on average, five points less than pupils with Norwegian background. The gender differential among immigrants is very similar to that of all pupils. The lower achievement of pupils with immigrant background is particularly strong in the lower end of the GP distribution. For example, 25 per cent of the non-western immigrant boys achieve less than 30 GP, see Table 5.1.

The unadjusted GP differentials displayed in this section represent (yet) another documentation of the importance of family background in predicting educational achievement. The differentials are substantial and there are numerous explanations for why they exist. An efficient policy to increase social mobility and reduce the impact of family background needs a thorough understanding of the underlying mechanisms. Such analyses go beyond the scope of this paper. The next section discusses the contribution of the different sets of family characteristics in explaining the overall GP variation.

[10] 'Immigrants' also include pupils born in Norway by two foreign-born parents.

Table 5.2. R^2-adjusted with different sets of individual and family background variables. Total, marginal and partial R^2

Pupil and family characteristics	Specification and controls	Total (R^2-adj)	Marginal (ΔR^2-adj)	Partial
I. Gender	Dummy-variable: Girl=1	0.06	0.06	0.06
II. Parental educational attainment	Educational attainment of mother and father (five groups, incl. missing), full interaction.	0.23	0.17	0.17
III. Family structure	Marital/co-habital status of the parents includes divorced and separated, number of full siblings and half-siblings. Birth quarter and order (among full siblings).	0.28	0.05	0.12
IV. Immigrant status	Country of origin (grouped). Age at immigration; [0-3), [3,6), [7,9), [9,11), [11,13) and 13+	0.29	0.01	0.03
V. Economic resources	Parental earnings during school years (age 6-15) and financial wealth (8th, 9th and 10th decile)	0.31	0.02	0.12
VI. Parental unemployment disability pension and social assistance	Parental unemployment, disability pension and social assistance by M and F during the last five years.	0.31	0.00	0.10

Note: Based on pooled estimations for the 2002 and 2003 cohorts, sample size is 106,494. 'Poor' parents are located in the lowest quintile (20%) in the parental earnings distribution and do not have financial wealth among the wealthiest 30%. Rich parents are in the highest earnings quintile and among the ten per cent with highest financial wealth.

5.2. What observed characteristics explain variation in grade points?

In this section we apply the regression framework given in equation (2) in Chapter 4 and estimate the effects of all background variables on individual GP simultaneously to obtain partial effects.[11]

A simple gender dummy specification 'explains' about 6 percent of the overall variation in grade points, see Table 5.2. The table displays the marginal contribution, defined as the change in the fraction explained variance (adjusted R^2) due to inclusion of a new set of controls, as well as the partial contribution (adjusted R^2- with these characteristics only).

The family characteristics are divided into five sets. Parental education is by far the most important set of characteristics.[12] The marginal R^2-adj is 0.17. Family structure also 'explains' a non-negligible part of the GP variation as the marginal contribution is 0.05 while the partial is as high as 0.12. *At the margin,* immigrant status does not contribute substantially, but the partial R^2-adj is 0.03. Given the relatively few immigrants, a partial contribution of 0.03 is not negligible.

Parental earnings and wealth do matter, even at the margin when parental education and family structure are included as controls (marginal R^2-adj equal to 0.02). The partial R^2-adj is the same as for family structure, 0.12.

Finally, information on parental unemployment, disability pension and social assistance received by mother and father during the last five years constitute the last set of family background characteristics. This set makes no marginal contribution to the explained variation, but the partial (when no other controls are included) is substantial, as partial R^2-adj = 0.10.

All in all, the total set of family characteristics 'explains' close to one third of the individual variation in school achievement, measured by the un-weighted average of marks across subjects by the end of compulsory schooling in Norway.

5.3. Conditional means ('archetypes') and marginal effects of family characteristics

In Table 5.3, the impact of the various family characteristics on the (conditional) mean GP is illustrated by means of archetypes. The idea is simple. For each archetype all variables are fixed to the sample average, except for those specified in the table. These conditional means are based on a regression on equation (2) with the most extensive set of family characteristics included.[13] By comparing archetypes in Table 5.3 we obtain estimates of the marginal effects of the varying characteristics. Thus, archetypes are not typical families, but they serve the purpose of describing the impact of a set of characteristics, holding all other variables constant.

The types A, B and C illustrate differences by parental education. Comparing the extremes, children of parents with an academic education where both have 17+ years of schooling have (on average) 10.5 higher GP than those with parents who did not attain education beyond the compulsory level. Note that this (conditional) differential is about two thirds of the unadjusted difference in Table 5.1.

Economic resources in the family matter, but the *marginal* effect of parental earnings and wealth is not very large. Comparing pupils from the 'poorest' quintile (type D) with those of rich parents with highest

[11] In section 5.4 we study the family background effects for some core subjects. In each subject, the individual mark is an average of exam results are marks awarded for classwork ('standpunkt'). Finally, in section 5.5 we apply the schools' own weighting schemes for subjects by using hours allocated to each subject throughout the lower secondary school.

[12] Remember that all the family characteristics are correlated so the marginal contribution of each set depends on the ordering.

[13] The coefficients used to attach importance to the various characteristics are collected from the 'pooled' 2002 and 2003 cohorts OLS regression based on the specification with all pupil and family characteristics included.

earnings and financial wealth (type E), the difference is about 3.9 GPs. Note that this differential is only a third of the unadjusted difference in Table 5.1, revealing that economics resources are highly correlated with other family characteristics.

The performance among Non-western immigrants who arrived in Norway before age 6 is (on average) very similar to that of other pupils with similar family background. Comparing type F and C, the immigrants actually perform better, but the standard errors are too large to claim significance. This suggests that the weaker performance of immigrants (Table 5.1) is explained by disadvantaged family background and the low marks achieved by immigrants arriving in Norway during school year ages.

Parental earnings and educational attainment are highly correlated. Pupils who have highly educated parents typically also benefit from spending their childhood and adolescence in a family without severe economic constraints. The extremely ‘advantaged’ and ‘disadvantaged’ family backgrounds are illustrated in types G and H, respectively. The differential is striking. Children of the richest and most educated parents scored about 17.8 GPs more than those with poor parents with low education who experienced severe unemployment and had to rely on social assistance.

A number of other individual and family structure characteristics do also influence pupil performance. Here we report selected results from the regression analysis, instead of constructing more archetypes (a complete list of estimates is not included). The ‘effects’ are all to be interpreted as marginal as they are estimated conditional on all other (family) characteristics.

First, the age at graduation matters. The pupils who finish the year they turn 15 is a selected group who have on average 1.6 (0.418) GPs more than the reference group who turn 16 during the first quarter of their final year.[14] Moreover, within the majority of pupils who turn 16 during their final year, pupils born during the first quarter perform better. Actually, among those who turn 16 during their final year, the GP is higher the older is the child at the time of school start (or completion). Relative to the reference group born during the first quarter, the differentials are -0.368 (0.070), -0.821 (0.066) and -1.364 (0.072) for the second, third and fourth quarter respectively.

Second, birth order and number of siblings affect school performance. They also interact. Consider first the average differentials across small families with one or two children. Relative to single children, first-borns do better, 0.479 (0.095), while the second-born has a disadvantage of -1.398 (0.096). Thus, first-borns have about 1.8 GPs more than their younger sibling. It also

Table 5.3. Pupil performance (GPs or ‘grunnskolepoeng’) by gender and family background. Predicted conditional mean. std.err.of the mean in parenthesis

‘Archetype’	Average GP	Deviations from the sample average
Girl	45.63 (0.06)	Gender
Boy	41.09 (0.06)	Gender
A. Highly educated parents	48.85 (0.17)	M(other) and F(ather) with a higher university degree (Master/Ph.D, 17+ years)
B. Parents with compulsory education	38.32 (0.18)	M and F with compulsory schooling only
C. Parents with upper secondary school	42.12 (0.07)	M and F with upper secondary school (10-12 years)
D. Rich parents	45.67 (0.11)	Parental earnings in the 5th quintile and wealth in the 10th decile
E. ‘Poor’ parents	41.77 (0.09)	Parental earnings in the 1st quintile and wealth in the 1st to 6th decile
F. Non-Western Immigrant*	42.83 (0.39)	Arrival in Norway at age 3-5, average non-western country of origin, mother and father 10-12 years of schooling (upper secondary level)
G. Rich and highly educated parents	51.20 (0.19)	Parental earnings in the 5th quintile, wealth in the 10th decile and both hold a higher university degree
H. ‘Poor’, unemployed parents with social assistance and low education	33.45 (0.41)	Parental earnings in the 1st quintile and wealth in the 1st to 6th decile, unemployed and social assistance recipient 3 years of the last 5 years and both have educational attainment at the compulsory level.
I. Non-western with ‘poor’, unemployed parents with social assistance and low education	34.16 (0.57)	Arrival in Norway at age 3-5, average non-western country of origin, mother and father 7-9 years of schooling. Parental earnings in the 1st quintile and wealth in the 1st to 6th decile, unemployed and social assistance recipient 3 years of the last 5 years and both have educational attainment at the compulsory level.

Note : *) Non-Western Immigrants include pupils born in Norway by foreign-born parents as well as first generation immigrants.

follows that single children on average do better that children with one brother/sister since (0.479-1.398)/2 < 0. The 'effect' of birth order is similar in large(r) families with three or more children. The first-borns score higher than their younger siblings, but also compared to single children. Again, relative to pupils without (full) siblings, the oldest have 0.912 (0.110) and 1.090 (0.185) higher GP, in families with three and four siblings respectively. The middle and last-born have lower GPs. In families with three, the estimates are -0.785 (0.112) and -1.607 (0.112) for the middle and last-born, respectively. The corresponding estimates for pupils in families of four are -0.95 (0.134) and -1.626 (0.185). Similar numbers are found in families with five or more siblings, but the estimates are less precise. To summarise, first-borns have an advantage. Pupil achievement is negatively related to family size, but the impact of the number of siblings is weak and non-linear. The marginal effect of family size is close to zero when the number of siblings exceeds two.

Third, pupils with biological parents who are not married have lower GPs. The 'effects' are statistically significant, but not large and considerably smaller than the unadjusted differentials in Table 5.1. Pupils with cohabiting parents score 1.14 (0.112) less and those with divorced parents 1.66 (0.084) less. The differentials for separated parents and the residual category are -1.83 (0.148) and -1.30 (0.105), respectively.

Fourth, pupils of young parents achieve lower GPs. In fact, GP appears to be a monotonic function of the age of the mother as well as the father's. To illustrate the importance, pupils with a teenage mother (at birth) score about 2.5 GPs less than pupils of mothers who were 35 years or more. The corresponding number for the age of the father is about 1.4 GP.

Finally, Table 5.3 revealed that the non-western immigrant archetype scores somewhat higher than the average pupil with Norwegian background and the same family characteristics. A few comments are appropriate here. First, this comparison is done by means of a regression analysis and not by means of matching pupils with observationally equivalent family characteristics. The impact of parental characteristics like education is assumed to be the same for all and thereby determined by the structure in the majority group. Family characeristics are important for ethnic minority children as well, but register data do not necessarily reflect family environment equally across cultures. One example is parental educational attainment for which information is missing for a substantial number of pupils with immigrant background. Moreover, the skills and competences that are relevant to children growing up in Norway and associated with parental educational attainment may also depend on whether the parents attended schools in Norway or abroad. Differences in quality of educational institutions and selective recruitment are examples of possible mechanisms. A study which focuses on why non-western immigrants are less successful in school, as shown in Table 5.1, would have to address these issues in detail. For our purpose, an exact assessment of the importance of ethnic minority background is less important. Second, the differentials across pupils with different ethnic/country background are substantial, up to 7-8 GPs on average. Third, pupils who arrive in Norway after age 9 achiew lower marks. For example, the 'marginal effect' of arriving with non-western immigrant background when aged 11-13 is -2.35 GP and very negative for those arriving after 13; -6.28. These differentials are far from unexpected. It would be highly surprising if 10 years in school in Norway did not give a huge advantage compared to children with 3 years or less. Needless to say, many immigrant children from non-western countries have experiences during their childhood and adolescence which have excluded learning opportunities and been detrimental to future learning capacity. Even children who went to school from age 6 in their country of origin and arrived in Norway during teen-age years, are unlikely to have acquired skills abroad which are completely transferable into a Norwegian context. However, even if these pupils are few, they do count in overall assessments of school performance.

5.4. Achievement and family background across subjects

This section reports on marks by family background in five selected subjects; Norwegian, mathematics, English, social studies and physical education.[15] In the three former core subjects, the score is the average of the marks awarded for classwork ('standpunkt') and that of the written examination (if taken). When we consider the subject specific marks below, without explicitly referring to classwork marks or exams, we look at the average of the two. Note the physical education is different form the other subjects since, according to official marking criteria, pupil ability should be taken into account when the mark is set. In other words, effort is likely to be more important than in other subjects.

Table 5.4 and 5.5 display how the sets of family characteristics contribute in explaining the variation in score by subject and classwork or examination marks. With the complete set of family background variables we 'explain' about 30 per cent of the variation in the total Norwegian score, about 25 per cent for English and mathematics, 23 per cent for social studies and significantly less for physical education (12 per cent). Thus, family background seems to be particularly important for achievement in the three core subjects.

[14] The number in parenthesis is the standard error of the estimated coefficient.

[15] There are two forms of the Norwegian language – Bokmål and Nynorsk. Norwegian-speaking pupils choose one form as their main language (primary form), and the other as their second language variant (secondary form). Norwegian here means the pupil's primary form.

Table 5.4. Family background and mark variation in Norwegian and mathematics. Adjusted R^2 by subject and sets of individual and family background variables.

	Norwegian			Mathematics		
	Total	Class work	Written exam	Total	Class work	Written exam
I. Gender	0.10	0.10	0.09	0.01	0.00	0.00
II. Parental educational attainment	0.24	0.23	0.20	0.17	0.16	0.17
III. Family structure	0.28	0.26	0.22	0.21	0.20	0.19
IV. Immigrant	0.28	0.26	0.22	0.21	0.20	0.20
V. Economic resources	0.29	0.27	0.23	0.23	0.22	0.22
VI. Parental unemployment, disability pension and social assistance	0.30	0.28	0.23	0.24	0.23	0.22

Table 5.5. Family background and mark variation in English, social studies and physical education. Adjusted R^2 by subject and sets of individual and family background variables

	English			Social studies			Physical education
	Total	Class work	Written Exam	Total	Class work	Oral exam	Class work
I. Gender	0.07	0.05	0.03	0.02	0.03	0.01	0.01
II. Parental education	0.21	0.18	0.16	0.17	0.17	0.14	0.07
III. Family structure	0.24	0.21	0.18	0.20	0.20	0.17	0.09
IV. Immigrant	0.24	0.21	0.18	0.21	0.21	0.17	0.09
V. Economic resource	0.25	0.22	0.19	0.22	0.22	0.18	0.11
VI. Parental unempl., disability pension and social assistance	0.25	0.22	0.19	0.23	0.23	0.19	0.12

Table 5.6. Pupil performance by subject, gender and family background. Predicted conditional mean. Std.err. of the mean in parenthesis

'Archetype'	Subject				
	Nor-wegian (primary form)	Mathe-matics	English	Social studies	Physical educ-ation
Girl	4.19 (0.01)	3.49 (0.01)	4.03 (0.01)	4.18 (0.01)	4.19 (0.01)
Boy	3.62 (0.01)	3.38 (0.01)	3.61 (0.01)	3.84 (0.01)	4.39 (0.01)
A. Highly educated parents	4.45 (0.02)	4.22 (0.02)	4.42 (0.02)	4.61 (0.01)	4.50 (0.02)
B. Parents with compulsory education	3.42 (0.02)	2.84 (0.02)	3.31 (0.02)	3.44 (0.02)	3,98 (0.02)
C. Parents with upper secondary school	3.77 (0.01)	3.28 (0.01)	3.68 (0.01)	3.87 (0.01)	4.26 (0.01)
D. Rich parents	4.11 (0.01)	3.75 (0.01)	3.99 (0.01)	4.28 (0.01)	4.53 (0.01)
E. 'Poor' parents	3.76 (0.01)	3.27 (0.01)	3.68 (0.01)	3.84 (0.01)	4.10 (0.01)
F. Non-Western immigrant	3.82 (0.04)	3.18 (0.05)	3.67 (0.04)	3.94 (0.05)	4.46 (0.04)
G. Rich and highly educated parents	4.65 (0.02)	4.53 (0.02)	4.59 (0.02)	4.88 (0.02)	4.74 (0.02)
H. 'Poor', unemployed parents with social assistance and low education	3.00 (0.04)	2.34 (0.05)	3.07 (0.05)	2.93 (0.05)	3.44 (0.05)
I. Non-western with 'poor', unemployed parents with social assistance and low education	3.05 (0.06)	2.24 (0.06)	3.06 (0.06)	3.00 (0.07)	3.65 (0.07)

Note: Total score, i.e. average of marks awarded for classwork and written/oral exam when applicable. Archetypes are explained in Table 5.3.

Family background explains a lower fraction of the variation in examination results, compared to what is explained of marks awarded for classwork. This illustrates that an achievement score based on a single test can be a noisy measures of cognitive ability. There are certainly other possible explanations. Teacher prejudice may operate in the favour of children from advantaged background and explain the higher explanatory power of family background when we look at marks awarded for classwork. Alternatively, as the exams do not give credit to effort its, the association between marks and family background can be different for classwork and examination marks.

The first rows of Table 5.4 and 5.5 also suggest that the gender differential is large in Norwegian, substantial in English, noticeable in social studies and smaller in mathematics and physical education. In Table 5.6, we report the predicted mean grade score by subject for the various archetypes introduced in Table 5.3.

Girls achieve, on average, a Norwegian score which is 0.57 higher than the boys. The gender differential is somewhat lower in English (0.42) and social studies (0.36). While boys are just behind in mathematics (0.1), they outperform the girls by 0.2 in physical education. These numbers illustrate that an overall measure of school performance can be sensitive to the weights given to each subject, see more discussion in section 5.5 below.

The impact of parental education is largest for mathematics where the difference between low educated (B) and highly educated (A) is about 1.4 points. The corresponding differential for physical education is only 0.5 and for the other three subjects about 1.1. The difference between children from rich (D) and poor families (E) is very similar across subjects, with a differential of 0.48 for mathematics and 0.31 for English as the two 'extremes'.

The achievement of non-western immigrants who arrived in Norway before age 6 is (on average) very similar to that of other pupils with similar family background. This similarity is found for all subjects. Comparing type E and C, immigrants actually do slightly better in three of the five subjects, but the standard errors are too large to claim significance. As explained in Chapter 3, the level of Norwegian language competence associated with a given mark, is different for an average immigrant and majority background pupil due to the inclusion of achievement in Norwegian as second language in the marks for a large proportion of immigrants. Thus, the performance of immigrants is systematically overvalued and this measurement error is of course particularly important when we focus on attainment in 'Norwegian' as in Table 5.6.[16]

The extremely 'advantaged' and 'disadvantaged' family backgrounds are illustrated by archetype G and H, respectively. Again, the differentials are striking; especially for mathematics (2.2) where the pupils with disadvantaged background score particularly low. The social studies differential is close to 2, Norwegian 1.7 and the English differential about 1.5. Again, physical education performance is less influenced by family background, but pupils with rich and highly educated parents (G) score 1.3 more than those born into to the other end of the family resource distribution (H).

As for the GP score, a number of other individual and family structure characteristics do also influence subject marks. Like in section 5.3 we discuss selected results from the regression analyses, without reporting the complete list of coefficients. Recall that these 'effects' are marginal in the sense that all other family characteristics are held constant. First, the age at graduation matters in all subjects. Within the majority of pupil who turn 16 during their final year, pupils born during the last quarter have a disadvantage of about 0.13 which is strikingly similar across subjects, compared to those born in January-March. There seems to be a linear age effect for all subjects with the second and third quarter in between. The minority of pupils who finish the year they turn 15 have do better in all subjects, except physical education. This selection effect is largest for mathematics and social studies (around 0.3), while the early starters score on average 0.15 higher in Norwegian and English.

Second, birth order and number of siblings do also affect subject marks. First-borns typically score higher than single children in all subjects, except English. For children with siblings, first-borns outperform schoolmates with older brothers or sisters. This holds for all subjects and the birth-order differentials are typically 0.3 or less. Performance is negatively related to family size in all subjects except physical education, but the impact of the number of siblings is weak and non-linear. The marginal effect of family size is close to zero when the number of siblings exceeds two. In physical education, however, the score is positively related to the number of siblings, but the 'effect' of another sibling is small.

Third, pupils with unmarried biological parents have lower scores in all subjects. The subject-specific differentials are statistically significant, but not large and around 0.1-0.2 for the various alternatives to married parents.

Fourth, pupils of young parents have lower achievement in all subjects, but the 'marginal effect' of parental age is lower for physical education than for other subjects. The differentials in other subjects are non-trivial. To illustrate the importance, pupils with a teenage mother (at birth) have a mathematics score which is 0.27 less than pupils of mothers who were 35 years or more, holding all other family characteristics constant. The corresponding numbers are similar for the other subjects, while the impact of the father's age is around 0.15.

Finally, the impact of age at arrival for non-western immigrants differs across subjects. While no disadvantage of late arrival is found for mathematics

[16] Note that we do not address the issue whether the practice with a lower level curriculum in Norwegian (Norsk 2.språk) is beneficial to immigrant pupils or not.

and physical education, those who arrive after aged nine score lower in social studies and Norwegian. Somewhat surprising perhaps, the disadvantage of late arrival is particularly large in English, where those who arrive at eleven or older score about 0.44 lower than other pupils. Recall, however, that the association between Norwegian language competence and age at arrival is likely to be distorted by Norwegian as second language. Presumably, a disproportionally large fraction of immigrants who arrived during school age years has Norwegian as their second language and thereby a separate curriculum.

5.5. One-dimensional measures of individual achievement: The importance of subject weights

We close this chapter with an assessment of how the aggregation of the various subject achievements affects the gender differential and 'impact' of family background. Any one-dimensional measure of achievement must aggregate across subjects. While the GP-measure attaches equal weights to the eleven subjects, the previous section revealed that differences across gender and family background characteristics are subject-specific. Presumably, people have different opinions on what kind of competencies which are more or less important. For example, numeric and quantitative skills during adolescence are more closely related to future labour market outcomes than are other competencies like e.g. language; see e.g. Paglin and Rufolo (1990), Doughtery (2002). At least, this suggests that mathematics qualifications are more important (for adult labour market outcomes) than ability to succeed in, say, music.

An alternative weighting procedure can build on the allocation of teaching hours given throughout the school years. Even with this principle, one has to decide on how to aggregate teaching hours across age levels as the composition of subjects varies by age. Here we report two alternatives, based on total hours for all ten years (1-10th grade) and the last three years (8-10th grade), respectively. The weights are reported in Table 5.7.

Although the weights depend on the grades included, a general pattern is that the 'teaching-hours-scheme' gives Norwegian and mathematics larger weights while those of music and home economics (heimkunnskap) are reduced[17]. Using the 8-10th grade hours, Christian knowledge and religious and ethical education (KRL) and arts and crafts (kunst og håndverk) also have weights below 1.1, see note Table 5.7.

17 Note that 'Norwegian' includes two marks, i.e. primary and secondary form, and is thereby given 'double weight' when using GP.

Table 5.7. Alternative weighting schemes based on teaching hours

Subject	Weights based on	
	Total subject hours 1-10th grade	Subject hours 8-10th grade
Norwegian (Norsk)	2.515	1.937
Mathematics (Matematikk)	1.716	1.522
Science and the Environment (Natur- og miljø)	0.917	1.245
English (Engelsk)	0.870	1.245
Social studies (Samfunnslære)	1.058	1.384
Physical education (Kroppsøving)	0.987	1.107
Christian knowledge and religious and ethical education (KRL)	0.964	0.899
Art and Crafts (Kunst og håndverk)	1.034	0.830
Music (Musikk)	0.611	0.415
Home Economics (Heimkunnskap)	0.280	0.415

Note: Teaching hours are taken from Lærerplan L97, page 81, and the subject weights (w_i) are calculated as follows; w_i = 11*(hours in subject i, grade j-k)/(total hours in the ten subjects, grade j-k). Multiplied by 11 to make it comparable with the grade point measure (GP) based on 11 subject (the ten above plus Norwegian secondary form).

Table 5.8. Pupil performance by weighting scheme, gender and family background. Predicted conditional mean. Std.err.of the mean in parenthesis.

Archetype	Weighting scheme		
	Total subject hours 1-10th grade	Subject hours 8-10th grade	Equal weights (GP) (from Table 5.3)
Girl	44.07 (0.07)	44.30 (0.07)	45.63 (0.06)
Boy	39.67 (0.09)	40.06 (0.08)	41.09 (0.06)
A. Highly educated parents	47.95 (0.20)	48.31 (0.20)	48.85 (0.17)
B. Parents with compulsory education	36.31 (0.21)	36.55 (0.21)	38.32 (0.18)
C. Parents with upper secondary school	40.57 (0.09)	40.87 (0.09)	42.12 (0.07)
D. Rich parents	44.44 (0.13)	44.81 (0.13)	45.67 (0.11)
E. 'Poor' parents	40.02 (0.12)	40.30 (0.12)	41.77 (0.09)
F. Non-Western immigrant	41.57 (0.45)	41.89 (0.45)	42.83 (0.39)
G. Rich and highly educated parents	50.57 (0.22)	50.98 (0.22)	51.20 (0.19)
H. 'Poor', unemployed parents with social assistance and low education	31.07 (0.47)	31.25 (0.48)	33.45 (0.41)
I. Non-western with 'poor', unemployed parents with social assistance and low education	32.07 (0.63)	32.26 (0.64)	34.16 (0.57)

The predicted achievement of each archetype, using the three alternative achievement measures, is displayed in Table 5.8. First, the two schemes based on teaching hours provide a very similar pattern. Hence, we focus on the subject hours 8-10th grade when we compare with the (equal weights) GP measure. Note that the numbers for the weighted measure are lower than for GP, reflecting that the former gives more weight to subjects with lower marks. For example, average marks in core subjects like mathematics are substantially lower than in physical education, see Table 5.6. Qualitatively, the differences across archetypes remain the same when we change to teaching hours weights. The magnitudes of the differentials are, however, somewhat different.

The gender differential is actually lower using 'teaching-hours-weights' (4.54 vs 4.24 GPs). The 'effect' of parental education is strengthened. Comparing the extremes, children of parents with an academic education where both have 17+ years of schooling (A) have (on average) 10.5 higher GP than those with parents who did not attain education beyond the compulsory level (B), while the differential is 11.8 using teaching hours weights.

Economic resources in the family also have somewhat higher impact using teaching hour weights since the difference between the rich quintile (D) and the 'poor' (E) is 4.5 compared to 3.9 GPs.

The marginal effect of immigrant background (F versus C, or I versus H) is basically independent of weighting scheme.

The combined effect of parental earnings and educational attainment can be studied by comparing type G and H, the extremely 'advantaged' and 'disadvantaged' family backgrounds, respectively. The differential is again somewhat higher using teaching hours weights; 19.7 versus 17.8 using GP.

5.6. Summary

This chapter confirms that a substantial part of the variation in achievement among 10th graders who complete compulsory education in Norway is explained by gender and family background characteristics available in administrative registers. Our main findings are:

- About one third of the variation in grade points (GP; un-weighted sum of marks in 11 subjects) is explained by gender and family characteristics
- Girls outperform boys by 4.6 GPs, on average.
- Individual GP is highly correlated with parental education, positively related to economic resources of the family, lower if parents have experienced unemployment or received social benefits, higher for pupils with united parents and also related to age (quarter of birth), number of siblings and birth order.
- Non-western immigrants who are born in Norway, or arrived before school start, achieve on average, approximately the same GPs as other pupils with comparable parental education and economic resources at home.
- Looking at the impact of gender and family background across five subjects we find that
  - The superior performance of girls is most prevalent in Norwegian, then English and less so in social studies. While boys are just behind in mathematics, they do better than girls in physical education. The impact of parental education is highest for mathematics and lowest for physical education.
  - Comparing extremely 'advantaged' and 'disadvantaged' pupils, the differential is highest (in descending order) for mathematics, Norwegian, English, social studies and physical education.
- Studying the importance of subject weights by comparing GP (i.e. equal weights) and an alternative measure based on teaching hour weights we find that the gender differential is basically the same. The 'effects' of parental education and economic resources are larger when subjects are weighted according to teaching hours.
- Finally, we emphasise that the 'effects' of family background characteristics are not proven to be causal.

6. Differences between schools: Persistent or random variation?

In this chapter we take a closer look at differences between schools with respect to *unadjusted* averages in total grade points (GP) and in single subjects illustrated by mathematics. The purpose of this exercise is to explore the extent to which such performance measures can be interpreted as effects of how schools operate or reflect random variation. Although we will argue that school results should be adjusted for differences in pupil composition with respect to family background, unadjusted mark distributions by subject at the school level have already been made public.[18]

School performance measured by average GP is a weighted average of individual achievement among the pupils. To simplify, achievement can be thought of as result of pupil composition, school quality and random variation. Pupil composition is important since family background is very important in explaining pupils' performance. Family background includes inherited abilities, learning both in the family and in earlier school years. 'School factors' is a label for a variety of characteristics that influence the learning environment and the processes that generate knowledge accumulation. Examples are resources, organization, management, cooperation, teacher qualifications (widely interpreted), etc. From a statistical perspective, both individual capacities of pupils and school factors are subject to random variation. Examples of such variation in pupil composition of pupils are family background, disabilities, learning during early school years etc. In short, the average capacity of the pupils changes from one cohort to the next. Examples of school factors prone to random variations are: teacher quality, sickness, flu epidemic, etc. However, school quality is naturally interpreted as the systematic component of school factors, those who remain fairly stable from one year to the next. Effects of radical reforms, organizational change etc, will however be hard to distinguish from 'random noise', at least with only a few years of data.

[18] In spite of the controversy on publication policy, the current practice is unlikely to be reversed. An assessment of the net value from publication is not addressed in this report.

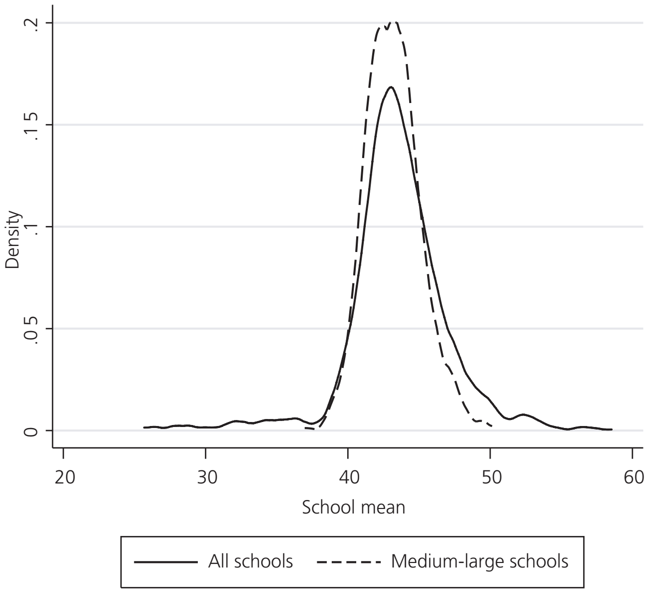
It is therefore important to know to what extent such measures reflect persistent differences between schools or noise and random variation. "School quality" is usually thought of as a "semi-structural" characteristic of a school, something that does not fluctuate too much from year to year: A measure that intends to reflect school quality should therefore show substantial stability over time.

A number of factors may contribute to "noise" in the school results; average school results may be affected by random variation that is uncorrelated between years or between different results within the same year. Such variation may be attributed to single pupils that perform particularly well or badly, specific teachers who have good or bad influence on the pupils, a flu epidemic on the day of exam etc.

Figure 6.1 shows the distribution of school mean GP for the years 2002 and 2003 combined, for all schools and for medium-large schools. Medium-large schools exclude small schools, that is schools with less than 31 pupils per grade per year (Given the maximum class size of 30 in grade 8-10 which applied up to 2003, this cut-off is equivalent to defining small schools as having one class per grade or several grades per class). Looking at the distribution for all schools, we see that there is considerable variation between schools in average results. The bulk of schools lie within an interval from 40 to 50 GPs, but the distribution has relatively thick "tails" of schools with very high and very low average GP.

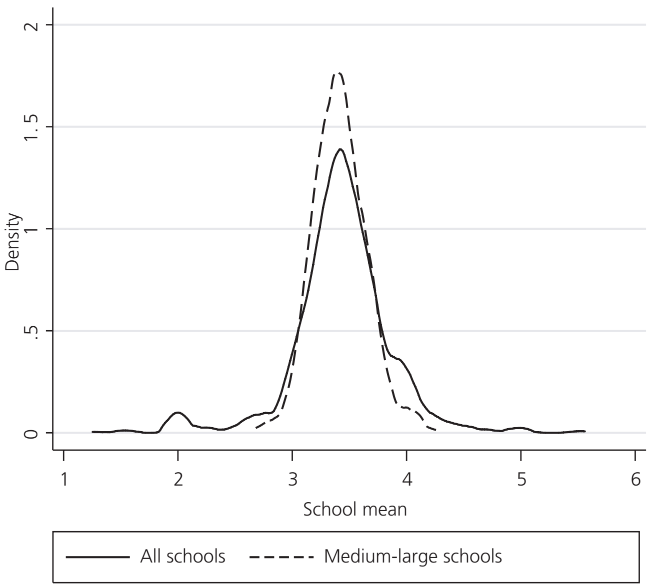
One of the most important results in theoretical statistics is that the standard deviation of a mean of observations equals the standard deviation of one single observation divided by the square root of the number of observations. This implies that the average GP of a school with 100 pupils in one grade will be five times more accurate (measured by the standard deviation) than the average GP of a school with 4 pupils. School results will therefore vary much more among small schools. In the absence of any systematic differences between schools with respect to pupil composition or "quality", the tails of the distributions will tend to be dominated by small schools.

Figure 6.1. Grade points distributions



Sample consists of 2002-03 cohorts combined,
(1253 observations, 578 medium-large)

Figure 6.2. School mean density. Unadjusted school mean mathematics marks

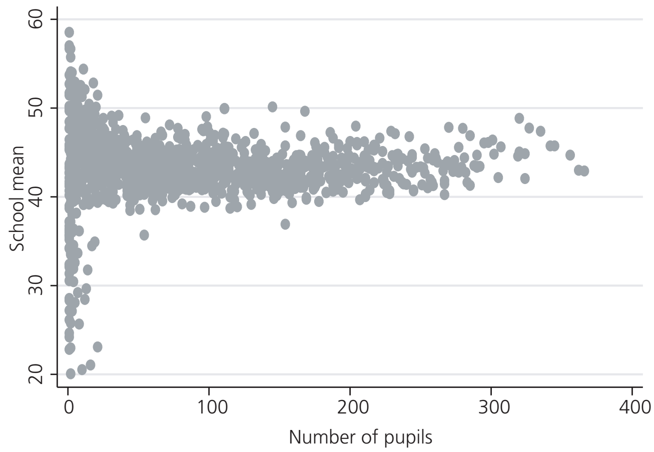


Sample consists of 2002-03 cohorts combined,
(1254 observations, 578 medium-large)

For medium-large schools we see that this distribution is much narrower, implying that small schools are more likely to have very high or very low averages scores than larger schools. The same pattern applies for single subjects, as illustrated by mathematics in Figure 6.2.

To illustrate the problem of random variation among small schools, Figure 6.3 plots average grade points against school size (e.g. the number of graduating pupils). We clearly see that there is much more

Figure 6.3. School mean and number of pupils. All schools



Sample consists of schools with school mean > 20,
2002-03 cohorts combined (1300 observations)

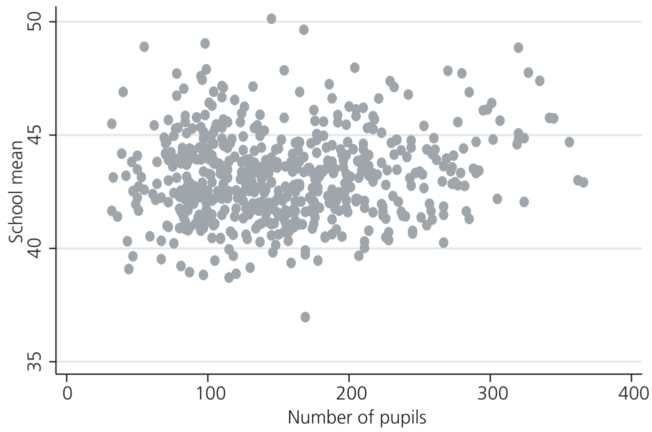
variation among small schools than among large schools, even if the figure combines two cohorts of pupils. The discussion above suggests that this for a large part may reflect random variation. In Figure 6.4 small schools are excluded. Here we see no clear tendency that variation is related to school size.

This shows that random variation - both pure "noise" and volatility of student composition with respect to family background - will give a tendency to rank small schools in the top and bottom of the distribution. This size effect does not reflect school quality. Exclusion of small schools, possibly combined with a pooling of graduation cohorts can be used as means to increase the number of observations behind each average and thereby reduce the influence from random 'noise'. However, the exclusion of small schools does not necessarily eliminate the problem of random variation. The larger the fraction of random variation of the variation in a school performance measure, the less persistent the performance measures will be over time. It is not entirely clear what should be counted as "low" or "high" degree of persistence in an absolute sense. Below, we look at persistence over time for different performance measures. In addition to graphical analysis, we use three different measures of correlation between performance measures, either between the same measure across different years, or different measures in the same period. We explain these measures by discussing Figure 6.5 and 6.6, which plot average school GP for 2002 and 2003, for all schools and large schools respectively. "R^2" (or R-squared) is a result from linearly regressing GP for 2003 on GP for 2002 and a constant term, and is the fraction of the variation between schools in GP for 2003 that can be explained by the variation between schools in GP for 2002. "Correlation" is the ordinary correlation coefficient, and "Rank correlation" is the correlation between the rankings of the two measures (ignoring the actual values of the measures). If GP for 2003 were an exact linear function of GP for 2002, all measures would equal 1. If it was an exact *nonlinear* monotonic function, the rank correlation would still equal 1, while the other measures would be less than one. In general, a lower degree of persistence of a measure or association between measures will result in lower values of R^2, correlation and rank correlation. There are also two lines in the figures. One is a 45-degree line. Schools with identical mean GP in the two years lie along this line. The other is the regression line for the two variables.

The figures reveal that schools that scored well in 2002 tended to score well also in 2003. This tendency is stronger if we look only at the large schools; all our correlation measures are higher for large schools than if we look at all schools. The regression line is also steeper in Figure 6.5. This strengthens the argument we made above, suggesting that school means are relatively more affected by random variation among small schools. Looking at the figures, we see that there are many schools with a substantial change in average GP from one year to the next. (The farther a point in the figure is away from the 45-degree line, the larger is the change in GP from 2002 to 2003).

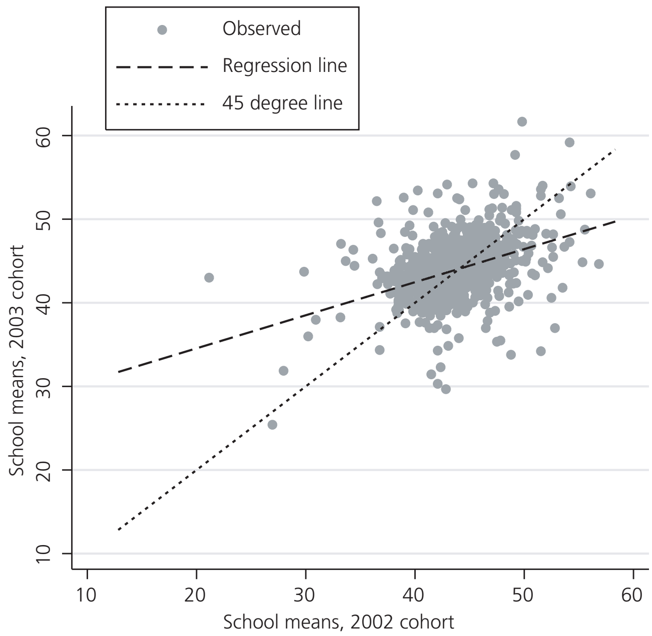
Another way to assess the degree of persistence is to look explicitly at how schools move up or down the GP distribution from one year to the next. To this end we ranked schools by their average GP, and divided them into ten groups of equal size (deciles) in each year. Looking at all schools, 36 percent stayed in the same decile or moved one decile from 2002 to 2003, 33 percent moved 2-4 deciles, while 31 percent moved 5 deciles or more. For large schools, the corresponding numbers are 40, 40 and 18 percent. It should be noted that differences between schools around the median are small, so a small change in a school's average GP may lead to a large change in its rank. Table 6.1 looks at persistence in the tails of the distribution. Given the position in the 2002 distribution, two persistence indicators are calculated: First the proportion which remains in the same group next year. These indicators are calculated for a division of schools into deciles (10 groups) and quintiles (5 groups). The second indicator is the proportion of school that cross the median; low-scoring school that move to the best half and high-scoring schools that move downwards. Table 6.1 shows that persistence is lower among the low scoring-schools. A school that scored low in 2002 is less likely to be in the lower end of the distribution in 2003 than the corresponding for a high-scoring school.

Figure 6.4. School mean and number of pupils. Medium-large schools



Sample consists of schools with school mean > 20, 2002-03 cohorts combined (581 observations)

Figure 6. 5. School means. 2002 versus 2003 cohort. All schools



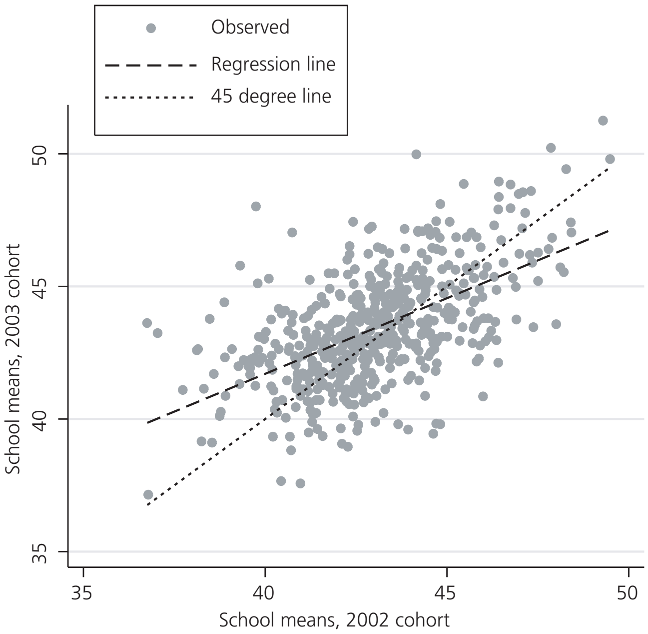
R2 = 0.1619, coeff. of correlation = 0.4024, rank corr. = 0.4256
Sample consists of 1042 observations

Table 6.1. Persistence in the top and bottom of the unadjusted school mean distribution. Number of schools, percent in parenthesis

2002:	# of schools	2003 In the same decile/quintile	Below median	Above median
Highest decile	55	23 (41.8)	7 (12.7)	48 (87.3)
Lowest decile	51	10 (19.6)	40 (78.4)	11 (21.6)
Highest quintile	110	56 (50.1)	21 (19.1)	89 (80.9)
Lowest quintile	102	42 (41.1)	80 (78.4)	22 (21.6)

Note: Percentiles defined by average 2002-2003. Medium-large schools.

Figure 6.6. School means. 2002 versus 2003 cohort. Medium-large schools



R2 = 0.3197, coeff. of correlation = 0.5654, rank corr. = 0.5530
Sample consists of 530 observations

The same tendency is revealed if we look at the persistence of average school marks in individual subjects across years. Table 6.2 reports persistence measures for individual subjects by school size. There appears to be more random variation among small schools. An interesting finding, however, is that the degree of persistence is lower for individual subjects than for GP. This could reflect the fact that GP itself an average over subjects, and thus reduces the impact of idiosyncratic subject-specific variation.

To summarize, it may be useful to think of differences between schools as consisting of three components: School quality, pupil composition and random variation. The analysis above suggests that the random variation component is substantial, particularly when the number of observations is low (small schools and individual subjects). To alleviate this problem, one should seek to increase the number of observations behind the performance measures, either by excluding small schools from the analysis, and/or by pooling several years of observations.

Table 6.2. Persistence of marks in selected subjects

Subject	Sample	Coefficient of correlation	Coefficient of rank correlation	R^2
Grade points	All schools	0.4024	0.4256	0.1619
	Medium-large schools	0.5654	0.5530	0.3197
Norwegian	All schools	0.2934	0.2905	0.0861
	Medium-large schools	0.3169	0.3107	0.1004
Mathematics	All schools	0.2617	0.2212	0.0685
	Medium-large schools	0.3481	0.3050	0.1212
English	All schools	0.2098	0.2061	0.0440
	Medium-large schools	0.3188	0.2570	0.1017
Social studies	All schools	0.3016	0.2960	0.0910
	Medium-large schools	0.3967	0.3775	0.1574
Physical education	All schools	0.2430	0.3370	0.0590
	Medium-large schools	0.3322	0.4607	0.1103

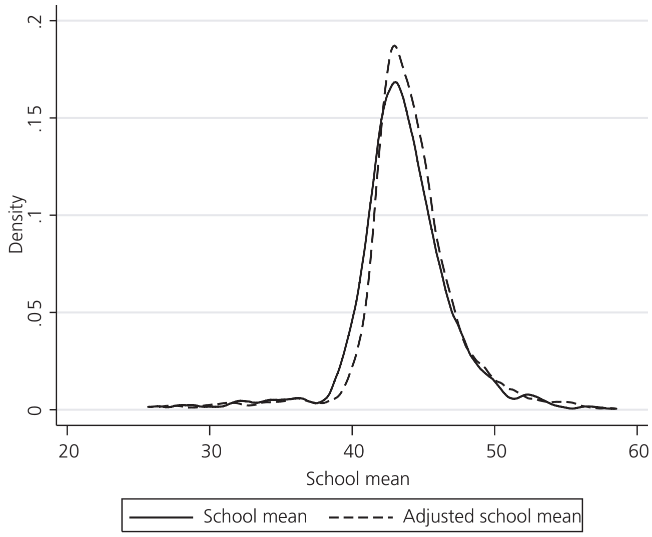
Note: 2002- and 2003 cohorts combined.

7. Towards the contribution of schools: Adjusting school means for family background

Pupils are not randomly assigned to schools, as residential location is the dominating determinant of school affiliation. Similar families tend to cluster in neighbourhoods and the pupil composition in the school reflects the socio-economic structure in the community in which the school is located. In previous chapters we have argued, both by referring to theoreticcal arguments, pointing to existing empirical findings and by our own empirical analysis, that family background characteristics are strongly related to individual school performance. Hence, it is of vital importance to adjust for differences in pupil composition if we want to put schools on an equal footing in an attempt to identify the contribution of schools to individual academic performance. In this chapter, we show how important such adjustments are in practice and how different types of family characteristics contribute to adjusting school results. Finally, we discuss the persistence of adjusted school means over time.

As described in Chapter 4, our school performance measure is estimated as the school fixed effect in a regression of individual grade points (or marks in single subjects) on a set of dummy variables indicating which school the pupil attended and our vector of family characteristics, cf. equation (2). The estimated school fixed effect can be interpreted as the average grade points for a school, conditional on family characteristics being equal to the sample mean. The adjusted school mean represents the hypothetical average grade points for the school if we 'replaced' their current set of pupils with a set of pupils with average family characteristics (and the grade points associated with this set of characteristics). In other words, the difference between adjusted and unadjusted means represents the impact on the observed average school mean which can be attributed to the fact that the school has a pupil composition that is different from the national average. For some schools, e.g. those in which the parents have low education, the adjustment effect is positive. Schools that have pupils with favourable family characteristics are adjusted downwards. In principle, correcting the school results from the contribution of pupil composition leaves us with an adjusted performance measure that is closer to the school's average contribution to individual performance[19]. However, this measure will still be influenced by random variation (see Chapter 6), and possibly also affected by unobserved pupil characteristics with effects on individual results which are unevenly distributed across schools.

Figure 7.1. Grade points distributions. School means. Unadjusted and adjusted. All schools



Sample consists of 2002-03 cohorts combined (1253 observations)

Figure 7.1 and 7.2 plot the densities of the unadjusted and adjusted school GP averages for the 2002 and 2003 cohorts combined for all schools and medium-large schools, respectively. They show that school results adjusted for family background characteristics have a narrower distribution than unadjusted results. Thus, placing schools on an equal footing by correcting for differences that can be attributed to differences in pupil composition reduces the variation between schools. Extreme results seem partly to be driven by pupil composition.

[19] Disentangling the contribution of schools and pupil characteristics requires that the allocation of pupils to schools is not too extreme. A stylized example may illustrate this. Suppose that gender is the only relevant pupil characteristic, and we have two schools, one for girls and one for boys. In this case it is not possible to identify school and gender effects separately. Note also that the adjusted school means reflects the average contribution and ignores the distinction that some schools can be favourable to specific types of pupils.

Figure 7.2.. Grade points distributions. School means. Unadjusted and adjusted. Medium-large schools

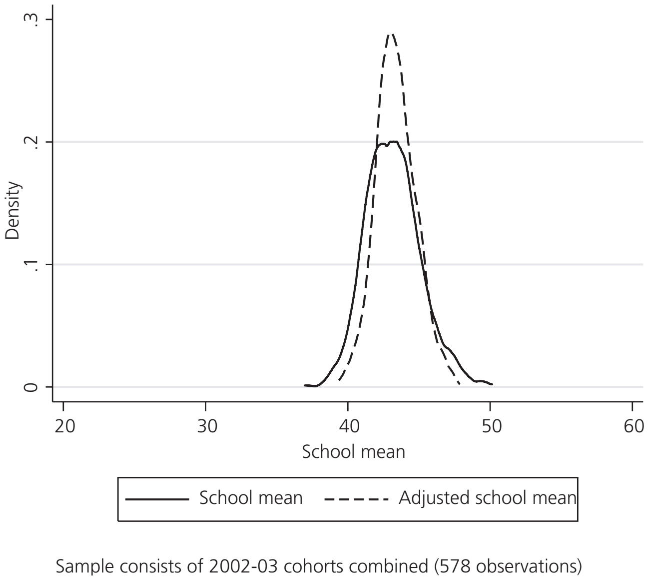
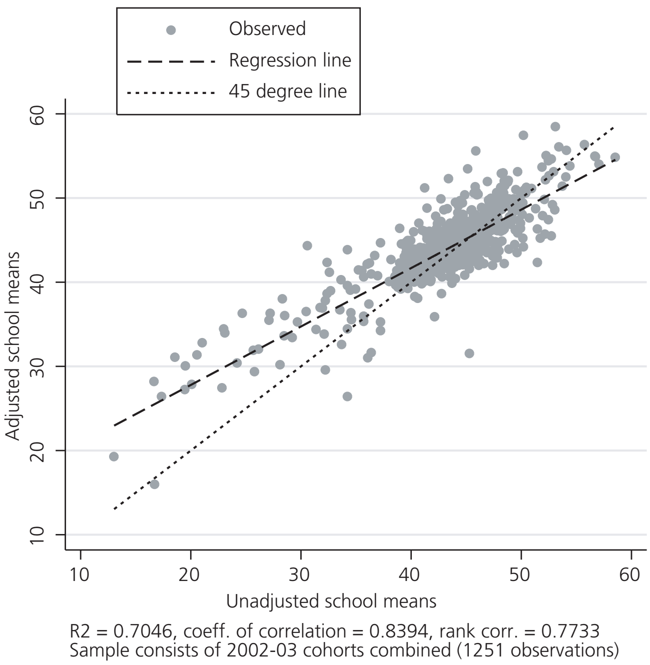


Figure 7.3. Unadjusted and adjusted school means. All schools



These distributions of unadjusted and adjusted GPs are informative, but they provide no insight into how the results for individual schools are affected by the adjustment. In spite of the declared intentions, the publication of any performance indicator with a school identifier, will ultimately lead to a ranking of schools, either by the publishing authority or by others such as newspapers, interest groups etc. The impact of adjustment on the ranking of schools is then a key question. If the unadjusted and adjusted rankings are very similar, one could naturally question the need for an adjustment at all since the two reflect the impact of schools equally well. On the other hand, if the rankings are very different, the unadjusted results are misleading measures of the contribution of schools.

Figure 7.4. Unadjusted and adjusted school means. Medium-large schools

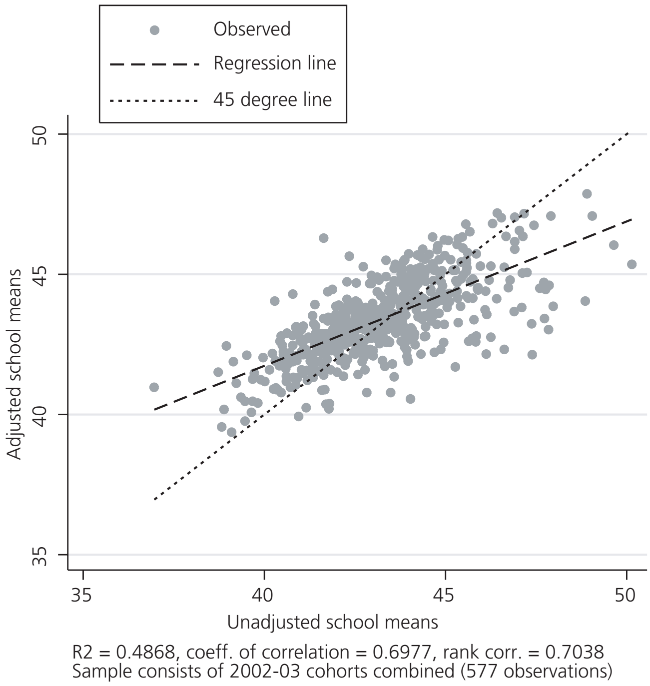


Figure 7.3 and 7.4 plot the unadjusted and adjusted school means for all schools and medium- and large-sized schools respectively. The sample is the 2002 and 2003 graduating cohorts combined. Each circle represents a school. The distance from the 45-degree line measures the adjustment effect. Schools above the line are adjusted upwards and vice versa. Thus, if the rankings were similar, all schools would be close to the 45-degree line. Both figures reveal a strong positive correlation between unadjusted and adjusted school means. Thus, removing the contribution of family characteristics, which we found explain around 30 percent of individual variation in GP, does not alter the main picture: Schools that performed well according to their unadjusted mean also tend to perform well according to adjusted means. The correlation measures between unadjusted and adjusted means are higher when we look at all schools than when we exclude schools with only one class per grade. This probably reflects what we pointed out in the previous chapter, that there is more random variation in school means among small schools. Thus random variation accounts for a larger share of total variation among small schools. Adjustment for pupil characteristics implies removing part of the systematic variation related to observed characteristics. Since this part constitutes a larger fraction of total variation among larger schools, the adjustment is larger and hence correlation between unadjusted and adjusted results is lower for medium- and large- sized groups.

However, even if the correlation is fairly high, a substantial number of schools move many steps up or down the ranking due to the adjustment. This is clearly illustrated by the figures as many schools are off the 45-degree line. In particular, many of the schools with the best and the worst performance according to the unadjusted results, are relatively far off the top and the bottom in the distribution of adjusted results. Table 7.1 shows how the ranking of schools by deciles change by adjustment, across the whole distribution. Many schools keep their ranking from the distribution of unadjusted results, but 36 percent (all schools) or 44 percent (medium-sized or large schools) move two deciles or more in the ranking. This shows that adjustment matters, in the sense that the pattern of schools that perform well or badly, change substantially.

Table 7.1. Effect of adjustment on ranking

Absolute change in rank, deciles	Distribution, percent	
	All schools	Medium/large schools
0	31.09	22.85
1	32.54	32.65
2	16.73	23.02
3-4	11.00	15.29
5+	8.63	6.19

Table 7.2. Adjustment effects in the top and bottom of the unadjusted school distribution

Unadjusted	# of schools	Adjusted		
		In the same decile/quintile	Below median	Above median
Highest decile	57	21 (36.8)	11 (19.3)	46 (80.7)
Lowest decile	58	28 (48.3)	56 (96.6)	2 (3.4)
Highest quintile	115	59 (51.3)	18 (15.7)	97 (84.3)
Lowest quintile	116	72 (62.1)	112 (96.6)	4 (3.4)

Note: 2002- and 2003 cohorts combined. Medium-large schools.

We now focus on the schools in the tails of the unadjusted school distribution. Where do they end up after adjustment? Does adjustment affect high- and low-performing schools differently? Table 7.2 shows that a*djustment has larger effects, for high-performing schools.* While only 36.8 remain in the 'top-10%' after adjustment, 48.3 % remain in the 'lowest 10%'. Second, of the schools belonging to the lowest decile or quintile, only 3.4% cross the median and belong to the upper half in the adjusted distribution after adjustment. However, close to 20 % of the pre-adjustment 'top-10%' drop below the median after adjustment.

We now decompose the adjustment of school results into contributions from the different sets of family characteristics discussed in Chapter 5. Table 7.3 and 7.4 show the contributions to the adjustment for the ten schools with the highest downward or upward adjustments among the medium- or large-sized schools. Table 7.3 shows that among schools with large downward adjustments, it is parents' education and economic resources that contributes most to the adjustment. These schools have pupils whose parents are particularly well educated and rich. Such factors affect school results, and the adjustment removes this effect. The contributions from the other sets of family characteristics are rather small. Among schools with a large upward adjustment, the pattern is somewhat different. Parental educational attainment is important for the adjustment here as well, but it is interesting to note that the effects of parental unemployment, disability pension and social assistance all have a quite large impact on the results for these schools. An important insight emerges from this exercise; even if some sets of family characteristics on the margin contribute little to explaining the *overall* individual variation in GP, they can still be important explanations for why some schools perform particularly well or badly according to the unadjusted school means.

Table 7.3. Decomposition of the largest downward adjustments. Contributions from different sets of pupil and family characteristics

Adjustment effect (AE)	Grade points before adjustment	Decile before adjustment	Grade points after adjustment	Decile after adjustment	Contribution to AE from:						Number of classes per grade
					Gender composition	Parental education attainment	Family structure	Immigrant status	Economic resources	Parental unemploym., disability pension and social assistance	
-5.39	47.39	10	42.00	2	0.25	-2.95	-1.02	-0.06	-1.16	-0.44	6
-4.96	47.84	10	42.88	4	0.12	-2.97	-0.85	0.01	-0.92	-0.36	5
-4.94	48.85	10	43.91	6	-0.08	-2.95	-0.71	-0.02	-0.79	-0.40	6
-4.90	50.13	10	45.23	8	-0.03	-2.77	-0.89	-0.08	-0.74	-0.39	4.5
-4.43	47.73	10	43.29	5	0.05	-2.33	-0.92	-0.07	-0.81	-0.35	5
-4.43	46.61	9	42.18	3	-0.10	-2.53	-0.64	-0.04	-0.77	-0.34	4
-4.28	47.38	10	43.11	4	0.04	-2.44	-0.76	-0.03	-0.72	-0.37	4.5
-4.27	47.97	10	43.70	5	0.18	-2.56	-0.87	-0.14	-0.66	-0.22	4
-4.11	46.13	9	42.02	2	-0.08	-2.26	-0.57	-0.02	-0.88	-0.31	4
-3.75	49.64	10	45.89	8	-0.13	-2.22	-0.68	-0.03	-0.34	-0.36	3

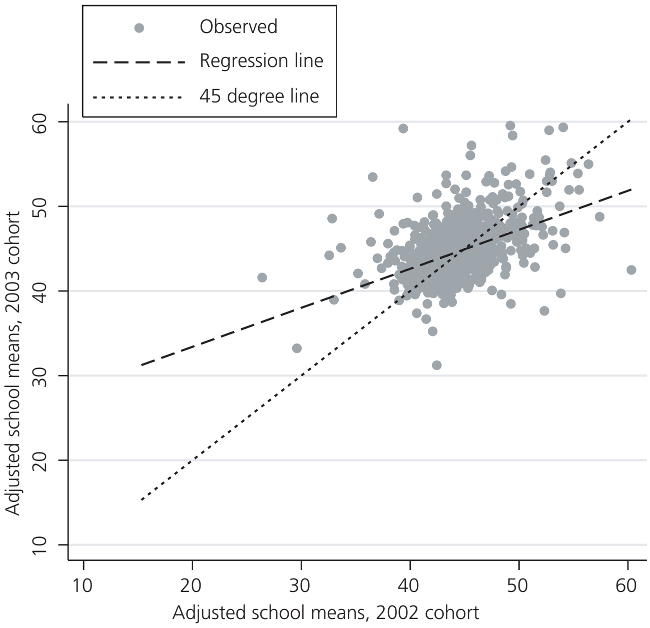
Note: 2002- and 2003 cohorts combined. Medium-large schools.

Table 7.4. Decomposition of the largest upward adjustments. Contributions from different sets of pupil and family characteristics

Adjustment effect (AE)	Grade points before adjustment	Decile before adjustment	Grade points after adjustment	Decile after adjustment	Contribution to AE from: Gender composition	Parental education attainment	Family structure	Immigrant status	Economic resources	Parental unemploym., disability pension and social assistance	Number of classes per grade
4.46	41.63	3	46.09	9	-0.13	1.92	0.47	0.12	1.07	1.01	5
3.84	36.96	1	40.80	1	-0.09	1.14	0.82	0.53	0.73	0.71	3.5
3.62	40.29	2	43.91	6	-0.30	1.37	0.74	0.08	0.76	0.96	3
3.36	40.78	2	44.14	6	0.08	0.96	0.53	0.03	0.75	1.00	4
3.34	38.95	1	42.29	3	0.03	1.44	0.68	0.12	0.58	0.49	2
3.13	42.34	4	45.47	8	0.33	1.28	0.38	-0.05	0.61	0.58	2
2.64	38.72	1	41.36	2	-0.07	1.40	0.46	0.02	0.41	0.42	2.5
2.57	39.15	1	41.72	2	0.02	1.31	0.09	0.11	0.66	0.37	3
2.46	41.75	3	44.21	6	0.07	1.18	0.48	0.06	0.42	0.26	2
2.44	41.87	3	44.31	6	0.43	0.52	0.26	0.67	0.43	0.12	3

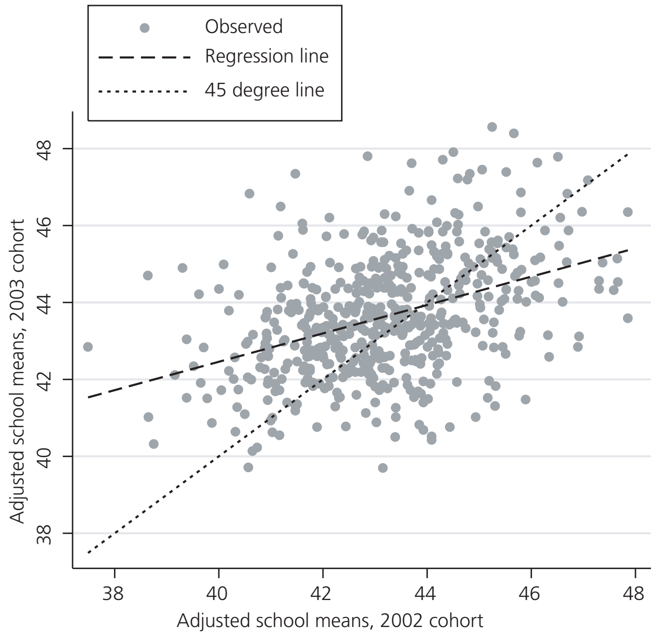
Note: 2002- and 2003 cohorts combined. Medium-large schools.

Figure 7.5. Adjusted school means. 2002 versus 2003 cohort. All schools



R2 = 0.2277, coeff. of correlation = 0.4772, rank corr. = 0.4786
Sample consists of 1042 observations

Figure 7.6. Adjusted school means. 2002 versus 2003 cohort. Medium-large schools



R2 = 0.1658, coeff. of correlation = 0.4072, rank corr. = 0.4034
Sample consists of 530 observations

One remaining issue is the stability over time of adjusted school means. If we believe that the adjusted school means reflect school quality, we would expect a substantial degree of persistence over time, as discussed in Chapter 6. To evaluate this, we estimated adjusted school means for the 2002 and 2003 cohorts separately. Figure 7.5 and 7.6 show that, as for unadjusted school means there is a clear tendency that schools that scored well in 2002 tended to do well in 2003 as well. The resemblance across years is weaker if we look only at the medium-sized and large schools; which is the opposite of what we found for the unadjusted school means in Chapter 6. Comparing persistence of adjusted means with unadjusted school means, the former it is higher if we look at all schools and lower if we exclude the small ones. A possible explanation for this is that year-to-year variation in pupil composition is larger for smaller schools. For larger schools, pupil composition is more or less constant over time. Removing family background composition as a source of variation takes away a stable component among the large schools and a more random factor among the small ones.

For specific subjects, the effects of adjusting for family background on the ranking of schools are more or less the same as for grade points. There are some variations between subjects, but the magnitude of the correlation measures correspond fairly well to differences between subjects regarding the impact of family background on individual marks, see Table 7.5.

Table 7.5. Correlations between unadjusted and adjusted school means, selected subjects. GP included for reference

Subject	Sample	Coefficient of correlation	Coefficient of rank correlation	R2
Grade points	All schools	0.8414	0.7739	0.7079
	Medium-sized and large schools	0.7469	0.7054	0.5578
Norwegian	All schools	0.8164	0.7710	0.6666
	Medium-sized and large schools	0.7149	0.7053	0.5110
Mathematics	All schools	0.8291	0.7828	0.6875
	Medium-sized and large schools	0.7202	0.7249	0.5187
English	All schools	0.8421	0.7928	0.7091
	Medium-sized and large schools	0.7479	0.7623	0.5594
Social studies	All schools	0.8316	0.7652	0.6915
	Medium-sized and large schools	0.7010	0.6771	0.4915
Physical education	All schools	0.9303	0.8985	0.8655
	Medium-sized and large schools	0.9186	0.8940	0.8439

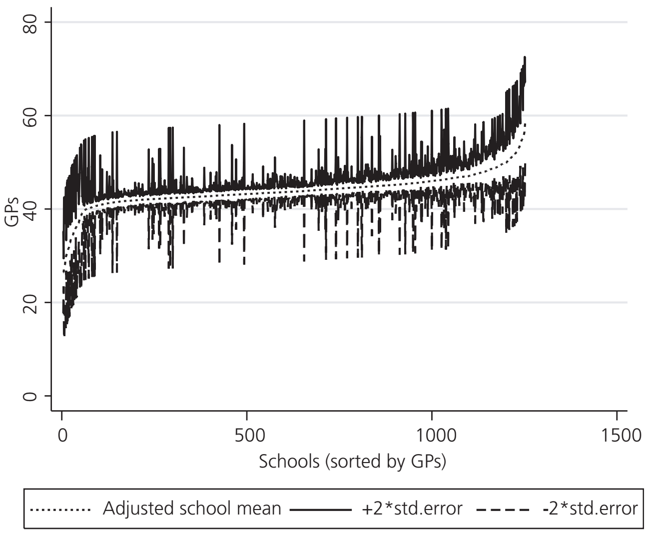
Note: 2002- and 2003 cohorts combined.

8. Schools differ – but are the differences statistically significant?

The two previous chapters showed that there are substantial differences between schools, both in terms of unadjusted average grade points and the school mean adjusted for the effects of family characteristics. However, we have also pointed out that there is statistical uncertainty associated with both measures. In this chapter, we assess to what extent our adjusted school means are affected by uncertainty, and with this in mind discuss how differences between schools should be interpreted.

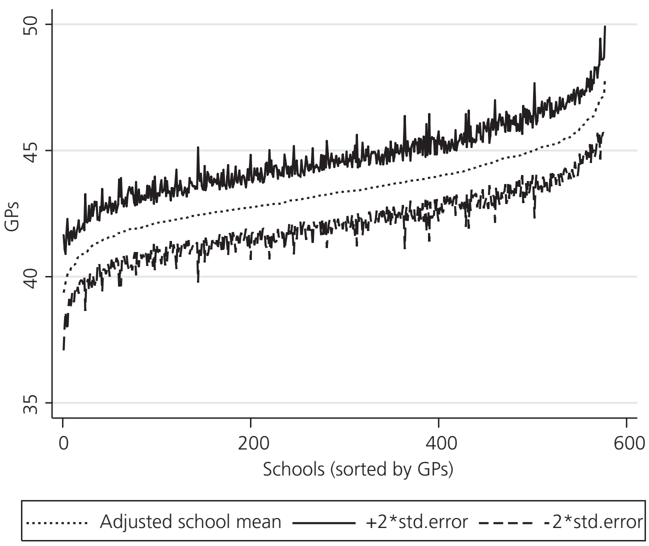
One may question why there should be uncertainty associated with a measure that covers (in principle) all pupils at a school. The observed average grade points for school A in year *t is* an exact measure of the average grade points *for school A in year* ***t***, since the typical sampling error is avoided. However, when we speak of uncertainty, the perspective extends beyond the observed unit. We want to know whether the observed distribution of school level performance, given the institutional constraints etc. in Norway reflect some more permanent variation in how schools operate, like 'school quality'. One way to look at this, as mentioned earlier, is that each individual mark or GP is a function of three components: School quality, pupil composition due to family background characteristics and random variation/unobserved factors at the individual level as well as at the school level. We *estimate* the contribution of family characteristics and there is some uncertainty arising from having estimates and not the true parameters. This spills over to the adjusted school means and contributes to the uncertainty of the adjusted school means. However, the major part of the uncertainty can be attributed to random variation ('noise') or variation in unobserved individual characteristics (which are not perfectly correlated with our observed family characteristics). This uncertainty can never be completely eliminated since each individual observation contains some random variation. As pointed out in Chapter 6, however, the more *independent* observations behind an adjusted school mean the lower the uncertainty. Hence, uncertainty will in general be smaller for larger schools. By pooling two cohorts, i.e. consider average school performance over a period of several years, the number of observations behind a mean will typically be doubled and thereby lower its uncertainty.

Figure 8.1. Adjusted school means



All schools, 2002-03 cohorts combined (1252 observations)

Figure 8.2. Adjusted school means



Medium-large schools, 2002-03 cohorts combined (577 observations)

The magnitude of this uncertainty must be studied empirically and will of course depend on the nature of the data at hand. Figure 8.1 and 8.2 present the estimated school fixed effects (adjusted school means) from equation (2), and with lines indicating plus/ minus two times the standard error of the school effects, roughly corresponding to a 95 percent confidence interval. Many of the adjusted school means are associated with substantial uncertainty. The uncertainty appears to be particularly large in the tails of the GP distribution. This is most apparent in Figure 8.1, indicating that uncertainty is largest among smaller schools, as discussed in Chapter 6. Figure 8.3 plots the standard errors of adjusted school means against school size, (the number of pupils in 10^{th} grade). We clearly see that the uncertainty is dominating for small schools, but declines asymptotically with the number of pupils.

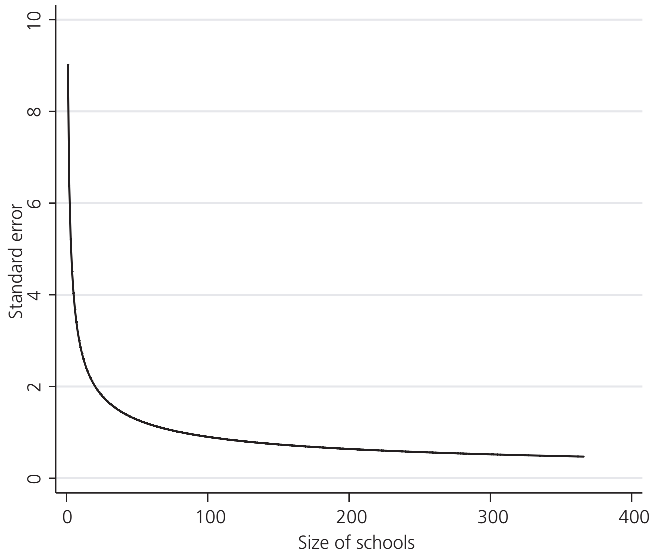
The substantial uncertainty associated with the adjusted school means indicates that many of the differences between schools are not statistically significant, even if they are large in numbers. Whether differences between schools are statistically significant or not has implications for how we should interpret the differences. It is therefore important to know how many of the differences that are significant and whether it is possible to establish rules of thumbs for how large a difference must be to be statistically significant. To examine this in detail, we looked at all possible pair-wise differences between schools (784,378) differences in the sample with all schools) and testes if the differences are statistically significant. Figure 8.4 and 8.5 present the results from this exercise. These figures have four curves. The cumulative density of differences in GP shows how large a fraction of differences in GP that is of a given magnitude or less. For example, Figure 8.4 shows that almost 80 percent of the between-schools differences are of five GP or less. The three other curves each indicate the fraction of differences in GP of a given size that are statistically significant at the 1, 5 and 10 percent level, respectively[20]. For example, Figure 8.4 shows that in the sample of all schools, around only 40 percent of differences in GP of size 10 are statistically significant at the 10 percent level, and that 90 percent of the differences are of this size or less.

The figures show that many differences between schools are not statistically significant. This is particularly striking, as expected, when small schools are included. Among medium- and large schools a less confusing picture emerges. Differences do not need to be extremely large to be statistically significant. Around 80 percent of differences of 2 GP are statistically significant at the 5 percent level. These amounts to a sizeable fraction of schools as around 30 percent of the differences are at least this large.

[20] The significance level is the probability that a null hypothesis is rejected, even if it is true.

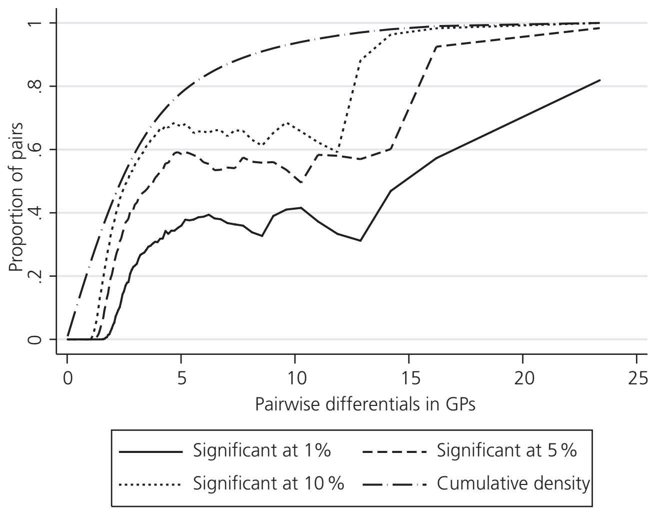
To summarize, the discussion in this chapter shows that there may be considerable uncertainty associated with the adjusted school means. This uncertainty is larger the fewer the number of observations per school. Restricting the sample to medium-sized and large schools and/or pool several years reduces the statistical uncertainty and thus a given difference between schools is more likely to be statistically significant. However, it is important to acknowledge that a large share of the differences between schools cannot be rejected to be generated by uncertainty. How large this share is depends on the choice of significance level. How to choose this level is not obvious, but the larger the consequences (of any kind) for schools of being classified as performing better or worse than others, the lower should be the level of significance

Figure 8.3. Standard error by size



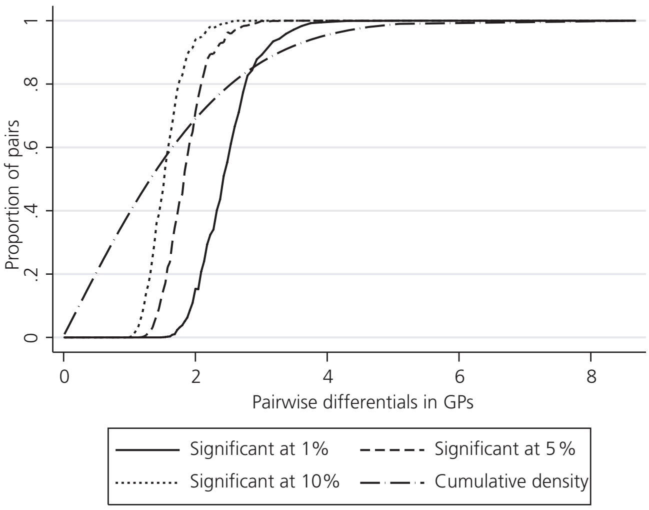
All schools, 2002-03 cohorts combined (1252 observations)

Figure 8.4. Significance of pairwise differentials. Adjusted school means against pairwise differentials



Sample consists of all schools,
2002-03 cohorts combined (784378 observations)

Figure 8.5. Significance of pairwise differentials. Adjusted school means against pairwise differentials



Sample consists of medium-large schools, 2002-03 cohorts combined (166753 observations)

9. Do differences in school resources matter?

This chapter investigates to what extent school differences, measured by adjusted school means, can be explained by teacher resources, qualifications and some other school characteristics. The empirical strategy is explained in chapter 4. While the main results are discussed in section 9.2, the following section defines our explanatory school variables in some detail and provides some descriptive statistics.

9.1. School characteristics[21]

Teacher hours, pupils and classes

This information is collected from the Compulsory School Information System ("Grunnskolens informasjonssystem" in Norwegian, henceforth GSI), over the years 1999-2002, covering grade 8-10 for the 2002- and 2003-cohorts. The basic inputs are yearly information at the school level on (i) the number of pupils in grade 8, 9 and 10 (ii) the number of classes by grade and (iii) the total hours of instruction ('årstimer til undervisning på ungdomstrinnet') for pupils in grade 8-10. Three measures of teaching resources are constructed, based on the following logic:[22] Instruction typically takes place within classes, but the number of teachers occupied with pupils belonging to a given class at a given point in time, varies across subjects, classes, grades and schools. If instruction is partly individual, larger classes all else equal reduce the teaching intensity allocated towards each pupil.[23] Teacher hours during a school year (*TH*), relative to the number of pupils (*P*), are an overall measure of instruction resources. This measure can be decomposed into two separate parts; teacher hours per class (*C*) and number of classes per pupil, i.e.

$$\frac{TH}{P} = \frac{TH}{C} \times \frac{C}{P}$$

or in natural logarithms

$$\ln\left(\frac{TH}{P}\right) = \ln\left(\frac{TH}{C}\right) - \ln\left(\frac{P}{C}\right),$$

where we have focused on the standard 'class-size' variable; pupils per class. In other words, higher total teaching resources relative to the number of pupils may be due to more teachers per class, either by more lessons or more teachers per lesson, and/or smaller classes. The three variables are constructed as the average of the three years prior to graduation (8^{th}-10^{th} grade) to reflect teacher input during the whole period the pupil attended lower secondary school, and not only during the final year.

Table 9.1 and Figure 9.1-9.4, illustrate a number of important features of variation in teacher input relative to the number of pupils. First, class size is (of course) closely related to cohort size and a huge part of the overall variation is found in schools with only one class. Among all schools, 25 percent of the schools have less than 12 pupils per class, see Table 9.1. This class size is far beyond any policy relevant variation. Class size below 15 is explained by sparsely populated areas and increasing decentralization of schools is hardly an issue. Reduction in standard class size towards a level below 20 is not likely. (Although recent changes in government regulations with respect to how the teaching is organized may make 'class size' an outdated description of teacher/pupil density, one can think of the number of teachers relative to pupils *per grade*)

Table 9.1. Class size and measures of teacher input. All schools and medium-large schools

		Percentiles				
	Mean	90 %	75 %	50 %	25 %	10 %
All schools (n=1133)						
Pupils / classes:	18.7	26.9	25.3	22	12	5
Teacher hours / classes:	1829	2156	1971	1807	1660	1524
Teacher hours / pupils:	173.3	380	142.6	85.7	73.2	66.0
Medium-large schools (n=559)						
Pupils / classes:	24.6	27.6	26.5	25.1	23.6	21.2
Teacher hours / classes:	1857	2114	1963	1824	1720	1621
Teacher hours / pupils:	76.3	90.7	82.6	74.5	68.2	62.9

[21] The reported results are based on the average of the school performance and characteristics for the 2002- and 2003-cohorts.
[22] To simplify exposition we ignore grades, although this distinction is accounted for in the construction of variables.
[23] A discussion of the mechanisms through which the pupil/teacher proportion matters is beyond the scope of this report.

Figure 9.1. Pupils / classes and number of pupils. All schools

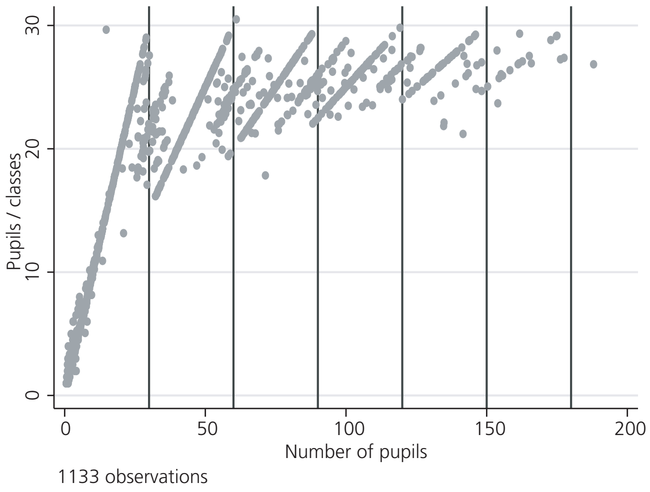


Figure 9.2.. Teacher hours / classes and number of pupils. All schools

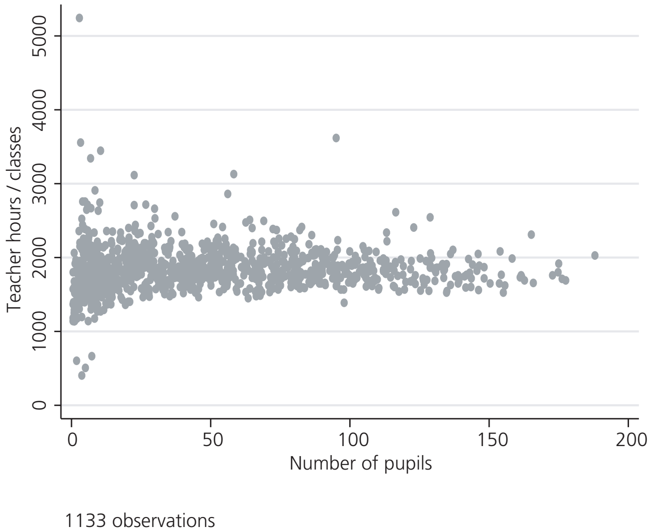


Figure 9.3. Teacher hours / pupils and number of pupils. All schools

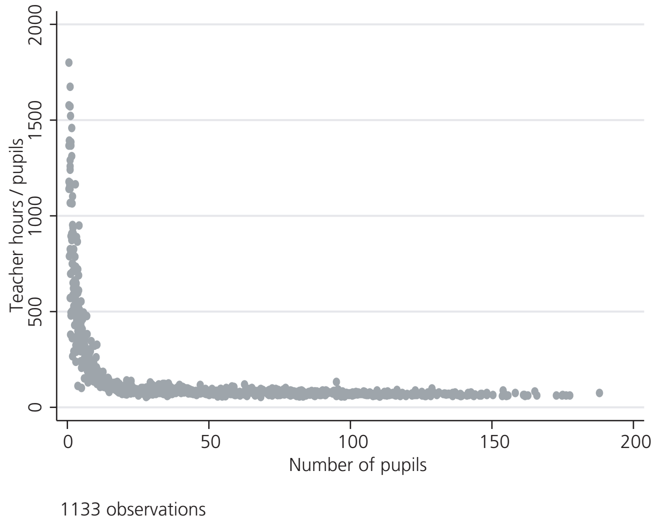
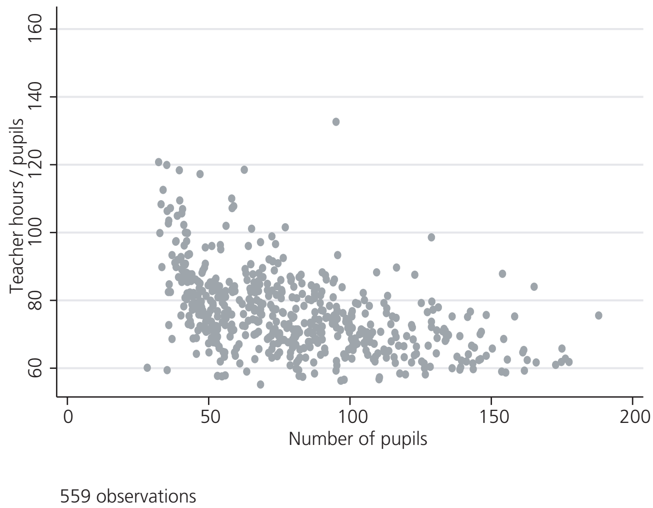


Figure 9.4. Teacher hours / pupils and number of pupils. Medium-large schools



The relationship in Figure 9.1 between average class size and number of pupils contains the familiar discontinuity around 30, 60, 90 etc. reflecting regulations on maximum class size of 30. Note the concentration of observations below the main trend. Recall that both measures are averaged over the three years the pupil attended lower secondary school, implying that new classes and closure across grades affect the three-year average.[24] Second, teacher hours per class are not closely related to school size. There are some indications that hours per class are increasing with size and the variation is larger among small schools, see Figure 9.2. However, the distribution for all schools and medium-large schools are fairly similar, see Table 9.1. Third, the variation in the overall measure 'teacher hours per pupil' is huge and completely dominated by few pupils per class in the sample of all schools; see Figure 9.3 and Table 9.1. If small schools are excluded, this combined measure 'makes sense', see Figure 9.4. There is considerable variation across medium-large schools, as those just behind the top 10 percent have close to 50 percent more teacher hours per pupil than the school at the first decile, see Table 9.1. Larger schools have less teacher hours per pupil, partly driven by the larger class size. Finally, there is a positive relationship between class-size and teacher hours per class (the coefficient of correlation in logs is 0.13 for medium-large schools and 0.35 for all schools). Thus, schools (or local authorities) seem to respond to larger classes by providing more teacher hours per class. Alternatively, they 'voluntarily' increase the number of classes and cut teacher hours towards the minimum at the same time.

[24] The relationship between pupils per class and the number of pupils, *by grade*, displays a much more distinct pattern, with discontinuities around 30, 60, 90 etc.

Ownership and grade coverage

GSI also provides information on whether the school is privately owned and whether the school is a combined primary and lower secondary school containing pupils from 1st -10th grade as opposed to 8th -10th grade lower secondary schools.

Many private schools do not report marks and are therefore not covered by this study. To illustrate, there are 73 private schools covered by GSI in 2000 with 10th graders (for the cohort graduating in 2001), but marks are available for only 48 of them. This is partly due to the evaluation methods used in Steiner-schools which deviates from the standard marking system and partly because many private schools do not report. The private schools are typically small and only four (!) are included in the main analysis of medium-large schools in section 9.2.

The majority of schools, 55 percent, are combined 1st - 10th grade among all schools. However, only 20 percent of the medium-large schools are combined schools.

Teacher qualifications

From the teacher statistics, we use detailed individual information on teachers to construct measures reflecting the level of formal qualifications, age profile, gender composition and turnover of the teacher staff. Unfortunately, teacher characteristics are based on the *total staff*, i.e. not matched to grades or classes in which the teachers work. Again, these variables are calculated as averages over the three years prior to graduation. Note that private schools are not included in the teacher statistics.

Teacher qualifications are based on educational attainment and classified by means of dummy variables if

(i) More than 10 percent of the teachers have university education at higher university level
(ii) None of the teachers have university education at higher university level
(iii) All teachers have formal teacher qualifications
(iv) More than 10 percent do not have formal teacher qualifications

The means of the four indicator variables are given in Table 9.2.

Turnover among teachers

The teacher statistics enables us to identify annual teacher flows in and out of each school, although not by grade affiliation. For each year we calculate the number of hirings (h_{jt}) and quits (s_{jt}) as well as the total number of teachers employed (n_{jt}). The churning rate ('gjennomstrekksrate') for year t in school j is defined as

$$churning\ rate_{jt} = \frac{h_{jt} + s_{jt} - |n_{jt} - n_{jt-1}|}{\frac{n_{jt} + n_{jt-1}}{2}}$$

Table 9.2. Teacher qualifications, turnover, age profile and gender

	Mean	Percentiles 90 %	75 %	50 %	25 %	10 %
Proportion written Norwegian exam:	0,2	0,51	0,47	0	0	0
1st-10th grade	0,2					
All teachers with formal qualifications	0,28					
> 10% of teachers without formal qualif.	0,02					
No teachers with higher tertiary education	0,12					
> 10% of teachers higher tertiary education	0,32					
Head teacher with higher tertiary education	0,11					
Proportion women:	0,57	0,71	0,63	0,56	0,5	0,43
Proportion age < 30:	0,13	0,23	0,17	0,12	0,07	0,03
Proportion age > 50:	0,38	0,53	0,46	0,37	0,3	0,24
Proportion same school last 3 years:	0,73	0,88	0,82	0,75	0,68	0,61
Churning rate	0,22	0,34	0,27	0,21	0,15	0,1

Thus, the churning rate is the sum of hirings and separations in excess of changes in the number of teachers employed, divided by the average number of teachers. The variable used below is the average churning rate over the three years in which the cohort attended lower secondary school. The average churning rate is 0.22, see Table 9.2. Our second turnover measure is based on the seniority profile of the teacher staff. For each year, we calculate the proportion of teachers who have been employed at the school all of the last three years, with an average across schools equal to 0.73, see Table 9.2.

Age profile and gender composition

The age and gender distributions of each school's teacher staff is measured by the fraction of teachers below age 30, the fraction above age 50 and the proportion of females. Descriptive statistics are given in Table 9.2.

9.2. Results

The analysis is based on the combined 2002- and 2003-cohorts for medium-large schools as explained in the previous section. A number of reasons motivate the focus on medium/large schools. First, as we have shown small schools are noisy when it comes to performance measures. Second, the resource variables are expected to be contaminated by measurement errors.They may include more than one grade in each class, and the turnover rates of teachers are volatile; if one or a few teachers leave the school a high turnover rate will follow. The same may be true for other characteristics of teachers and pupil composition, since small changes will induce large variation in measured variables. Measurement error leads to attenuation of the results and imprecise estimates. These problems are less important when we restrict the sample to medium-large schools. We do exclude a large fraction of the schools, but a small proportion of the pupils. Finally, much of the variation is resource use comes from small schools in sparsely populated areas. This variation is far outside the interval relevant for policy.

In light of the extensive evidence in previous chapters on the importance of adjusting for differences in family background, we use adjusted grade point averages as the school performance measure. Table 9.3 presents the estimates of the components in the school production function (equation (4) in chapter 4)[25].

Column I reports the estimates when school resources are included as the composite measure 'teacher hours per pupil'. The main result is that teacher input matters, but not very much. The estimated coefficient is 1.26 with a standard error of 0.417, which means that the effect is significantly different from zero. Statistical significance does not necessarily imply importance. Two simple indicators illustrate this. First, all the variables included only explain a small proportion of the variation in performance across schools as the adjusted R-squared is 0.04. However, even if we explain a limited proportion of the overall variation, the effect of a single variable can be substantial. This is not the case here, as shown by the predicted GP differential between a school at the 90th and the 10th percentile in distribution of 'teacher hours per pupil'. The school with more teacher hours per pupil is estimated to gain an average increase in GP of about 0.5, see bottom of Table 9.3. More precisely, the difference in average grade points for schools at the higher end of the resource use distribution as compared to those at the lower end is 0.46 grade points. Although the adjusted grade point distribution of schools is fairly narrow as shown in Figure 7.2, the estimated effect suggests that teacher input measured by hours of instruction is of limited, but not of negligible, importance.

Column II in Table 9.3 offers an answer to the question: Are teaching hours per class more important than pupils per class, or are the two equally important? Teacher hours per class has the expected positive sign, indicating that more teacher hours per class improve the average performance of pupils. The effect is significantly different from zero. Even class size has the expected negative sign. Larger classes have detrimental effect on school performance, but the effect is not very precise and not significantly different from zero (at 5 percent level). The split in column II does not improve the explanatory power of the model and the total contribution of all school variables is unchanged (as the adjusted R-squared is still only 0.04). The predicted 90-10 percentile differential is about 0.40 GP for teacher hours per class which means that the school which use more teaching resources scores about 1/3 GP better. The predicted differential for class size is 0.24, which means that a school on the 10th percentile with 21.2 – on average – achieve one quarter of a GP more than a school on the 90th percentile with 27.4 pupils per class. The confidence interval for this prediction is wide and does cover zero.

Table 9.3. Teacher input, school characteristics and school performance

Dependent variable: Adjusted GP school average	Model with	
Variables:	I. Teacher hours per pupil	II. Teacher hours per class & Pupils per class
ln (Teacher hours/Pupils)	1.26	
	(0.417)	
ln (Teacher hours/Classes)		1.53
		(0.525)
ln (Pupils/Classes)		-0.921
		(0.581)
All teachers with formal qualifications	0.072	0.083
	(0.125)	(0.126)
> 10% of teachers without formal qualif.	-0.021	-0.023
	(0.412)	(0.412)
No teachers higher tertiary education	-0.155	-0.131
	(0.202)	(0.204)
> 10% of teachers with higher tertiary education	0.077	0.074
	(0.126)	(0.126)
Head teacher higher tertiary education	-0.296	-0.297
	(0.180)	(0.181)
Proportion female teachers	-1.292	-1.316
	(0.650)	(0.651)
Proportion teachers < 30 years old	-0.003	0.047
	(0.971)	(0.920)
Proportion teachers > 50 years old	0.342	0.355
	(0.572)	(0.573)
Proportion teachers at school last 3 years	-0.902	-0.902
	(0.500)	(0.495)
Churning rate ('gjennomtrekksrate')	-0.360	-0.372
	(0.721)	(0.722)
1-10th grade	0.632	0.652
	(0.186)	(0.188)
Proportion of pupils with exam in Norwegian, written form	-0.181	-0.177
	(0.233)	(0.233)
Constant	43.328	43.32
	(0.057)	(0.057)
Adjusted R-squared	0.04	0.04
Number of schools	548	548
Predicted GP differential between teaching resources at the 90 and the 10 percentiles. 95 % confidence interval in parenthesis		
ln (Teacher hours/Pupils)	0.4580 [0.1608, 0.7551]	
ln (Teacher hours/Classes)		0.4028 [0.1311, 0.6745]
ln (Pupils/Classes)		-0.2429 [-0.5451, 0.0592]

Note: Medium and large schools (> 30 pupils). Private schools excluded. 2002- and 2003-cohorts.

[25] To account for the fact that the school performance measures are estimated with uncertainty which varies systematically with school size, equation (4) is estimated using weighted least squares, using the square root of the number of pupils as weights.

Table 9.4. Teacher input, school ownership and performance

Variables:	Model with dependent variable I. Unadjusted GP school average	II. Adjusted GP school average
Ln (Teacher hours/Pupils)	-2.459 (0.567)	1.195 (0.397)
Private school (=1)	5.990 (1.067)	3.158 (0.747)
1-10th grade	0.158 (0.237)	0.425 (0.166)
Proportion of pupils with exam in Norwegian, written form	-0.158 (0.324)	-0.275 (0.227)
Constant	43.24 (0.080)	43.35 (0.056)
Adjusted R-squared	0.08	0.06
Number of schools	559	559

Note: Medium and large schools (> 30 pupils). Private schools included. 2002- and 2003-cohorts.

How important are our teacher characteristics in explaining school outcomes? Teachers' qualifications, using categorical measures of the composition of the educational background of the schools' teachers, have the expected signs. Higher qualifications improve school performance and low qualified staff reduces school performance. However, the variables are imprecisely measured and the estimates are not significantly different from zero. A head teacher with a long university education seems to be a disadvantage, but the effect is very imprecise and far from significant. If we turn to the age composition of the staff of teachers, we notice no effects whatsoever. The share of female teachers has a negative effect on performance, but again the estimate is highly imprecise and not significantly different from zero. Increased turnover of teachers also has a negative, but insignificant effect on school performance. In sum, teacher qualifications in terms of educational background, experience as teachers, school seniority, and gender composition, do not appear to be important determinants of differences in school performance among Norwegian lower secondary schools.

Although Norway is characterized by a dominant public school sector, there is a small number of private schools at the lower secondary education level. Table 9.4 provides some highly indicative evidence on the relative performance of private schools. The two columns report estimates of the difference, using the unadjusted average GP and adjusted average GP, respectively. The table reveals that the private schools score about 6 GPs higher than the average public school, when we do not control for family background of the pupils. This is a huge difference in performance. However, when we control for differences in family background, the difference drops by one half. Still, this is a substantial difference in light of the narrow distribution of adjusted GP in Norway across schools. There are several reasons why this evidence must be considered highly indicative. First, only four (!) private schools are included in the sample, for reasons discussed in section 9.2. We have no indication that these are representative. Second, we do not know anything about other school characteristics. Finally, one might suspect that private school pupils – on average – have higher learning capabilities, even if we condition on observed family characteristics. Much more is to be said about this issue, but a thorough analysis requires more detailed data on inputs and results in private schools.

The evidence so far suggests that teaching resources matter, but the average effect of teacher input used at school during the three final years of compulsory schooling, is limited and not precisely determined. As widely discussed in the literature, and briefly touched upon in chapter 4, the identification of causal effects of school resources face a number of problems. One is related to how resources are allocated across schools. These processes are crucial as we use the actual, implemented, non-experimental, distribution of (teaching) resources to identify the parameters of interest. There are numerous administrative regulations and specific measures within the school sector which link the allocation of resources to a particular school and the ability or learning 'capacity' of the pupils enrolled. This can take place at the individual level through specific decisions ('enkelt-vedtak') or via 'objective' allocation rules that link resources to the composition of pupils. In short, compensatory resource allocation along these lines implies that estimates of the effects of teacher input on school performance are likely to be biased downwards. For instance, if schools respond to enrolment of weaker pupils by providing more teaching hours and smaller classes for this group of pupils, we will underestimate the effect of resource use in schools.

We present a simple test of how this type of allocation may be affecting our results. The logic of the test is as follows. We know that observed family background is very important in explaining pupils' performance. It is highly plausible that our family characteristics correlate with factors that both release extra resources and vary systematically with the capacity of the pupil. Imagine we look at the (deliberately miss-specified model represented by the) relationship between unadjusted school performance and teacher input. By comparing the effect of school resources for unadjusted and adjusted school average grade points, we implicitly test whether compensating allocation of school resources creates a substantial bias. If compensatory practice is widespread and important, we should get a much weaker estimated effect of school resources when not controlling for pupils' family background. Possibly, the estimate will have the opposite sign as compensatory use will bias the effect of resources downwards. In Table 9.5 we report the results for unadjusted average grade points. The explanatory variables are identical to those in Table 9.3.

There is a noticeable difference between the results with and without controlling for family background. Concentrating first on the results for the composite resource measure in column I in Table 9.5, the resource variable changes sign from positive to negative and the ‘perverse’ effect of ‘teacher hours per pupil’ is even highly significant. The same happens with the specification reported in column II when we split ‘teacher hours per pupil’ into ‘pupils per class’ and ‘teacher hours per class’. This shows that resources are allocated in a compensatory way by spending more resources on expectedly weaker pupils or pupils with a less favourable family background. Moreover, it is possible that this type of allocation affects our estimates, even when we adjust for a wide range of family background characteristics. In other words, this exercise suggests that even the effects of teacher input reported in Table 9.3 may be biased towards zero.

Turning to the effects of teachers’ qualifications using unadjusted GP, an interesting result is that teachers’ qualifications now seem to matter, in contrast to the almost non-existing effects when we do control for family background. Teacher qualifications are positively related to school performance. However, these school characteristics are not proven to have causal effects, but highly qualified teachers are overrepresented in schools with well-performing pupils. Or, schools with less qualified teachers have pupils with lower marks. This evidence suggests that assortative matching of teachers and pupils is taking place at the lower secondary level in Norway[26]. The same picture appears when we look at the impact of the proportion of women in the school. In the specification with unadjusted GP, the effect of a high female share of teachers on school performance is highly positive. This is the opposite of the result for adjusted GP. The interpretation is the same; female teachers are overrepresented in schools with pupils with advantaged family background. Or, male teachers tend to work in schools with pupils who have – on average – lower achievement.

We focus on *average* effects, ignoring a number of issues and perspectives that are highly relevant for policy making. We would like to emphasize that the data at hand offer numerous possibilities to explore questions like e.g.; Are teacher resources more important for the performance of low-scoring pupils, or do clever students gain more? Is the effect of marginal increases in school resources the same throughout the resource distribution? Or is it possible to identify thresholds, where performance is hit particularly hard if resources drop beyond this level? Do school resources affect the performance dispersion within school? Is there a trade-off between raising the average performance and reduce the within-school variation? What kinds of teacher hour inputs are important? How does the compensating resource allocation really work and what are the implications for empirical studies of school quality?

A number of school production studies have been published recently on Norwegian survey data, Bonesrønning (2003), (2004a), (2004b). The matched register data used in this study offer a number of new avenues for research, but there are also important limitations as information on school/teacher practices are not included.

Table 9.5. Unadjusted school means. Teacher input, school characteristics and school

Dependent variable: GP school average (unadjusted) Variables:	Model with: Teacher hours per pupil	Teacher hours per class &Pupils per class
ln (Teacher hours/Pupils)	-1.674 (0.573)	
ln (Teacher hours/Classes)		-1.148 (0.721)
ln (Pupils/Classes)		2.346 (0.800)
All teachers with formal qualifications	0.514 (0.176)	0.535 (0.172)
> 10% of teachers without formal qualif.	-0.505 (0.566)	-0.508 (0.566)
No teachers with higher tertiary education	-0.521 (0.277)	-0.475 (0.280)
> 10% of teachers higher tertiary education	0.515 (0.173)	0.507 (0.173)
Head teacher with higher tertiary education	0.178 (0.247)	0.176 (0.247)
Proportion female teachers	3.647 (0.893)	3.600 (0.893)
Proportion teachers < 30 years old	-1.407 (1.260)	-1.308 (1.265)
Proportion teachers > 50 years old	0.815 (0.786)	0.840 (0.786)
Proportion teachers at school last 3 years	-0.675 (0.679)	-0.675 (0.679)
Churning rate (‘gjennomtrekksrate’)	-0.011 (0.991)	-0.037 (0.991)
1-10th grade	0.053 (0.256)	0.088 (0.257)
Proportion of pupils with exam in Norwegian, written form	-0.155 (0.320)	-0.146 (0.320)
Constant	43.17 (0.078)	43.16 (0.078)
Adjusted R-squared	0.11	0.11
Number of schools	548	548
Predicted GP differential between teaching resources at the 90 and the 10 percentiles. 95 % confidence interval in parenthesis		
ln (Teacher hours/Pupils)	-0.6073 [-1.0154, -0.1992]	
ln (Teacher hours/Classes)		-0.3022 [-0.6751, 0.0707]
ln (Pupils/Classes)		0.6188 [0.2040, 1.0335]

Note: Medium and large schools (> 30 pupils). Private schools excluded. 2002- and 2003-cohorts.

[26] Bonesrønning, Falch and Strøm (2004) also find evidence of teacher sorting with respect to the share of immigrant pupils in Norwegian schools.

10. Summary and conclusions

The aim of this report is to provide background information on how to construct informative performance indicators, based on pupils' mark achievements, for schools at the Norwegian lower secondary education level ("Ungdomsskolen").

Though not perfectly, marks measure the cognitive competence level of pupils. The pupils' level of cognitive skills is an important starting point for assessing the contribution of schools to pupil learning. Schools may also be important providers of non-cognitive skills, but this issue extends beyond the scope of our report.

Our point of departure is the present situation. Average marks by school *are* being published and schools *are* being compared. The question of whether marks should be published or not is therefore outside the scope of this report. The question is rather: *How* should it be done? Is it possible, given the available data, to construct indicators that are closer to reflect the contributions of schools than the unadjusted differences in marks? Should such adjusted indicators be published?

It is widely accepted that unadjusted differences in marks by school do not necessarily reflect differences in school quality. Such differences may exist because of (i) the composition of pupils (of which some aspects are measurable), (ii) random variation (both at the individual and school level) and (iii) difference in school quality. The relative contributions of (i) - (iii) in explaining differences across Norwegian lower secondary schools are unknown. Available data have so far not been utilized to disentangle the different factors. Partial analyses based on survey data, everyday experience of parents and teachers and a pile of evidence on family background and educational outcomes, all suggest that (i) and (ii) are important, but there is little evidence of how important they are. Hence, it was unknown to what extent unadjusted differences in marks by school reflect differences in school quality.

Our data covers all pupils who completed compulsory education in Norway (10^{th} grade in the lower secondary school) for the years 2001 to 2003. We use three main types of data in our analysis containing information about pupil marks by subject, individual characteristics and family background, as well as schools. Most of the data are taken from various administrative registers. The use of common identifiers of individuals and schools across registers facilitates matching of different registers and enables us to construct a dataset that is very rich in both individual and school characteristics.

For the 2002- and 2003 cohorts, individual marks by subject are available, both marks awarded for classwork ("standpunkt") and for examinations. Average marks for eleven subjects are used to calculate an aggregated achievement measure, labelled total grade points (GP; "grunnskolepoeng"). We also present analyses of some core subjects.

Random variation - both pure "noise" and volatility of student composition with respect to family background - will give a tendency to rank small schools in the top and bottom of the distribution. This 'size effect' does not reflect school quality but follows from theoretical statistics telling us that the standard deviation of a mean of observations equals the standard deviation of one single observation divided by the square root of the number of observations. Exclusion of small schools, possibly combined with a pooling of graduation cohorts, can be used as means to increase the number of observations behind each average and thereby reduce the influence from random 'noise'. However, exclusion of small schools does not necessarily eliminate the problem of random variation.

Different correlation measures show that high-scoring schools in 2002 also tended to score well in 2003. We ranked schools by their average GP, divided them into ten groups of equal size (deciles) in each year, and found that about 40 percent stayed in the same decile or moved one decile from 2002 to 2003, 40 percent moved 2-4 deciles, while 18 percent moved 5 deciles or more (looking at schools with more than 30 pupils per grade).

It should be noted that differences between schools around the median are small, so a small change in a school's average GP may lead to a large change in its rank. Persistence is found to be lower among the low-scoring schools. A school that had a low score in 2002 is less likely to be in the lower end of the distribution in 2003 than the equivalent for a high-scoring school.

A substantial part, about one third, of the variation in individual achievement (pupil's GP) is explained by gender and family background characteristics available in administrative registers. Family is by far the most important factor explaining pupils' cognitive performance. Our main findings are: Girls outperform boys by 4.6 GP, individual GP is highly correlated with parental education, positively related to economic resources of the family, lower if parents have experienced unemployment or received social benefits, higher for pupils with united parents and is also related to quarter of birth, number of siblings and birth order. Non-western immigrants achieve, on average, approximately the same GPs as other pupils with comparable parental education and economic resources.

Looking at the impact of gender and family background *across* five subjects we find that the superior performance of girls is most prevalent in Norwegian, then English and less so in social studies. While boys are just behind in mathematics, they do better than girls in physical education. The impact of parental education is highest for mathematics and lowest for physical education.

Studying the importance of subject weights by comparing GP (i.e. equal weights) and an alternative sum based on teaching hour weights, we find that the gender differential is basically the same and that the 'effects' of parental education and economic resources are larger when subjects are weighted according to teaching hours.

It is of vital importance to adjust for differences in pupil composition if we want to put schools on an equal footing in an attempt to identify the contribution of schools to individual achievement. We show that school performance (average GP) adjusted for family background characteristics has a narrower distribution than unadjusted results. Extreme results seem partly to be driven by pupil composition.

Publication of any performance indicator identifying schools, will ultimately lead to a ranking of schools, either by the publishing authority or by others such as newspapers, interest groups etc. The impact of adjustment on the ranking of schools is then a key question. Schools that performed well according to unadjusted means also tend to perform well according to adjusted means. However, even if the correlation is fairly high, a substantial number of schools move many steps up or down the ranking due to the adjustment. Many schools keep their ranking from the distribution of unadjusted results, but 44 percent (medium-large schools) move two deciles or more in the ranking. Although many schools are clustered in the middle of the distribution implying that a small change in average GP for a school leads to movements in the distribution, this shows that adjustment matters. The pattern of schools' performance, i.e. which do well or less well changes substantially. We also find that the adjustment has largest effects for high-performing schools. While only 36.8 remain in the 'top-10%' after adjustment, 48.3 % of the lowest performing schools remain in the 'lowest 10%'.

For the schools with the highest downward or upward adjustments among the medium-large schools, we decompose the adjustment of school results into contributions from the different sets of family characteristics. Among schools with large downward adjustments, it is parents' education and economic resources that contribute most to the adjustment. These schools have pupils whose parents are particularly well educated and rich. Even among schools with a large upward adjustment, parental educational attainment is important for the adjustment, but it is interesting to note that the effects of parental unemployment, disability pension and social assistance are substantial for these schools. An important insight emerges from this exercise; even if some sets of family characteristics on the margin contribute little to explaining the *overall* individual variation in GP, they can still be important explanations for why some schools perform particularly well or badly according to the unadjusted school means.

The substantial uncertainty associated with the adjusted school means, implies that many of the differences between schools are not statistically significant, even if they are large in numbers. This is particularly striking, as expected, when small schools are included. Among medium-large schools differences do not need to be extremely large to be statistically significant. Around 80 percent of differences of 2 GP are statistically significant at the 5 percent level. This amounts to a sizeable fraction of schools as around 30 percent of the differences are at least this large.

Finally, we pick up on the distribution of adjusted school means and investigate the extent to which school differences can be explained by teacher resources, qualifications and some other school characteristics. Three measures of teaching resources are constructed based on the idea that instruction typically takes place in classes. However, the number of teachers occupied with pupils belonging to a given class at a given point in time, varies across subjects, classes, grades and schools. If instruction is partly individual larger classes all else equal reduce the teaching intensity allocated towards each pupil. Higher total teaching resources relative to the number of pupils may be due to more teachers per class, either by more lessons or more teachers per lesson, and/or smaller classes. The three variables are constructed as the average of the three years prior to graduation (8-10th grade) by means of data from the Compulsory School Information System ("Grunnskolens informasjonssystem" (GSI)), to reflect the teacher input during the whole period for which the pupil attended lower secondary school, and not only during the final year. The characteristic 'Teacher hours per class' is not closely related to school size. There is some indication that hours per class increases with size and the variation is larger among small schools. However, in the sample of alls schools the variation in the overall measure 'teacher hours per pupil' is huge and completely dominated by few pupils per class in small schools. If small schools are excluded, this combined measure 'makes sense'.

GSI also provides information on whether the school is privately owned and whether the school is a combined primary and lower secondary school containing pupils from 1-10th grade as opposed 8-10th grade lower secondary schools. Very few private schools are included in the sample of medium-large schools since many private schools do not report marks and the majority of these schools are small. From the teacher statistics, we use detailed individual information on teachers to construct measures reflecting the level of formal qualifications, age profile, gender composition and turnover of the teacher staff.

The analysis is based on the combined 2002- and 2003-cohorts for medium-large schools. We exclude small schools because they are noisy when it comes to performance measures, the resource variables are expected to be contaminated by measurement errors, and because much of the variation in resource measures among small schools are generated by residential location and found in sparsely populated areas far outside the variation relevant for policy. In light of the extensive evidence on the importance of family background adjustment in previous chapters, we use the adjusted grade point averages as the school performance measure.

The main result is that teacher input matters, but not very much. The effect of 'teacher hours per pupil' is significantly positive. The school with more teacher hours per pupil is estimated to gain an average increase in GP of about 0.5. More precisely, the difference in average grade points for schools using many teacher hours (90 percentile) as compared to those at the low end (10 percentile) is 0.46 grade points. When we split into teacher hours per class and pupils per class both measures appear to matter independently. Teaching hours per class has the expected positive sign, indicating that more teacher hours per class improve the average performance of pupils. The effect is significantly different from zero. Even class size has the expected negative sign. Larger classes have detrimental effect on school performance, but the effect is not very precise and not significantly different from zero (at 5 percent level). The predicted 90-10 percentile differential is about 0.40 GP for teacher hours per class which means that the school which use more teaching resources scores a bit less than half a GP better. The predicted differential for class size is -0.24 which means that a school on the 10th percentile with 21.2 pupils per class on average achieves one quarter of a GP more than a school on the 90th percentile with 27.4 pupils per class. The confidence interval for this prediction is fairly wide and does cover zero.
Differences between schools with respect to teachers' qualifications in terms of educational background, experience as teachers, school seniority, and gender composition, do not appear to be important determinants of differences school performance among lower secondary schools in Norway.
The private schools score about 6 GPs higher than the average public school, but the difference drops to about 3 when we control for family background. There are several reasons why this evidence must be considered indicative: Only four (!) private schools are included in the sample, and we have no indication on whether these are representative. We have no information on other school characteristics and one might suspect that private school pupils – on average – have higher learning capabilities, even if we condition on observed family characteristics.

We also discuss the implications for identification of teacher input effects when compensatory resource allocation takes place across schools. A number of administrative regulations and specific measures within the school sector link the allocation of resources to a particular school and the ability or learning capacity of the pupils enrolled. This can take place at the individual level through specific decisions ('enkelt-vedtak') or via 'objective' allocation rules that link resources to the composition of pupils. In short, compensatory resource allocation along these lines implies that estimates of the causal effects of teacher input on school performance are likely to be biased downwards.

We present a simple test of how this type of allocation may be affecting our results. Imagine that we look at the (deliberately misspecified model represented by the) relationship between unadjusted school performance and teacher input. By comparing the effect of school resources for unadjusted and adjusted school average grade points, we implicitly test whether compensating allocation of school resources creates a substantial bias. If compensatory practice is widespread and important, we should get a much weaker effect of school resources when not controlling for pupils' family background and possibly the opposite sign as compensatory use will bias the effect of resources downwards.

We find a noticeable difference between the results with and without controlling for family background. The resource variable 'teacher hours per pupil' shifts sign from positive to negative and this 'perverse' effect is even highly significant. The same happens when we split 'teacher hours per pupil' into 'pupils per class' and 'teacher hours per class'. This shows that resources are allocated in a compensatory way by spending more resources on expectedly weaker pupils or pupils with a less favourable family background. Moreover, it is possible that this type of allocation affects our estimates, even when we adjust for a wide range of family background characteristics.

Turning to the effects of teachers' qualifications using unadjusted GP, teachers' qualifications now seem to matter, in contrast to the non-existing effects when we do control for family background. This means that while teacher qualifications are positively related to school performance, they are not found to have causal effects. Highly qualified teachers are overrepresented in schools with well-performing pupils. This evidence suggests that assortative matching of teachers and pupils is taking place at the lower secondary level in Norway. The same picture appears when we look at the impact of the proportion of women in the school. In the specification with unadjusted GP, the effect of a high female share of teachers is highly positive on school performance. This is the opposite of the result for adjusted GP. The interpretation is the same; female teachers are overrepresented in schools with pupils with advantaged family background. Or, male teachers tend to work in schools with pupils who have – on average – lower achievement.

References

Aamodt, P. O. (1982). *Utdanning og sosial bakgrunn*. Samfunnsøkonomiske studier nr. 51, Statistisk sentralbyrå.

Altonji, J.G. and Dunn, T. A. (1996). "Using siblings to estimate the effect of school quality on wages." *Review of Economics and Statistics*, Vol. 78(4), pp. 665-671.

Angrist, J. and Lavy, V. (1999). "Using Maimonides' rule to estimate the effect of class size on student achievement", *Quarterly Journal of Economics*, May 1999, pp. 533-575.

Arnesen, C.Å. (2003). *Grunnskolekarakterer våren 2003*. NIFU skriftserie 32/2003.

Bakken, A. (2003). *Minoritetsspråklig ungdom i skolen.* NOVA-rapport 15/03.

Baydar, N. and Brooks-Gunn, J. (1991). "Effects of maternal employment and child-care arrangements on preschoolers cognitive and behavioural outcomes: evidence from the children of the NLSY." *Development Psychology,* VOl. 27, pp. 932-945.

Betts, J. R. (1996). "Do school resources matter only for older workers?", *Review of Economics and Statistics*, Vol. 78, pp. 638-652.

Black, S., Devereux, P. and Salvanes, K. G. (2003). Why the apple doesn't fall far: Understanding Intergenerational Transmission of Human Capital", NBER working paper no 10066.

Bonesrønning, H. (2003), Class Size Effects on Student Achievement in Norway: Patterns and Explanations. *Southern Economic Journal*, 69(4), 952-965.

Bonesrønning, H. (2004a) The determinants of parental effort in education production: do parents respond to changes in class size? *Economics of Education Review* 23 (2004) 1-9.

Bonesrønning, H. (2004b) Can Effective Teacher Behavior be Identified? *Economics of Education Review*, Vol 23, No.3, 237-247

Bonesrønning, H., T. Falch og B. Strøm (2004): Teacher sorting, teacher quality, and student composition: Evidence from Norway, *European Economic Review*, forthcoming

Browning, M. and Heinesen, E. (2003) "Class size, teacher hours and educational attainment", CAM Working paper 2003-15, University of Copenhagen.

Card, D. and Krueger, A. (1992). "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, vol. 100 (1), pp. 1-40.

Cameron, S. and Heckman, J. J. (1998). "Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males." *Journal of Political Economy*, Vol. 106(2), pp. 262-333.

Carneiro, P. and Heckman, J. J. (2002). "The evidence on credit constraint in post-secondary schooling." *Economic Journal*, Vol. 112, pp. 705-734.

Coleman, J. S. et al. (1966). *Equality of Educational Opportunity*. Washington DC: US GPO.

Dearden, L., Ferri, J. and Meghir, C. (2002)."The effect of school quality on educational attainment and wages", *Review of Economics and Statistics*, Vol. 84, pp. 1-20.

Dougherty (2002) "Numeracy, literacy and earnings: evidence from the National Longitudinal Survey of Youth", *Economics of Education Review* 21 (2003), 511-521.

Dustmann, C. Rajan, N. and van Soest, A. (2003). "Class size, education and wages." *Economic Journal*, Vol. 113(1), pp. F99-F120.

Goldstein, H. and Spiegelhalter, D. J. (1996). “League tables and their limitations: Statistical issues in comparison of institutional performance.” *Journal of Royal Statistical Society Series A*, Part 3, pp. 385-443.

Hanushek, E. A. (1996). “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, E. A. (2003). “The Failure of Input-Based Schooling Policies”, *Economic Journal, Vol*. 113 (February), pp. F64-F98.

Hanushek, E. A. and Raymond, M. E. (2004). “The effect of school accountability systems on the level and distribution of student achievement.” Forthcoming *European Economic Review*.

Hanushek, E. A., Kain, J.F. and Rivkin, S. (1998). “Teachers, schools and achievements.” NBER working paper no. 6691.

Haveman, R. and B. Wolfe (1995): “The Determinants of Children’s Attainments: A Review of Methods and Findings”, *Journal of Economic Literature*, Vol XXXIII, No.4, 1829-78.

Heckman, J. J, Ichimura, H. and Todd, P. (1998): "Matching as an Econometric Evaluation Estimator," Review of Economic Studies, Vol. 65 (2). 261-94.

Hernes, G. and Knudsen, K. (1976). Utdanning og ulikhet. Levekårsundersøkelsen Oslo: NOU 1976: 26.

Hoxby, C. (2000). “Peer effects in the classroom”, NBER working paper No. 7867.

Kane, T. J. and Staiger, D. O. (2001). “Improving school accountability systems.” NBER working paper no 8156.

Kane, T. J. and Staiger, D. O. (2002).”The promise and pitfalls of using imprecise school accountability systems”, *Journal of Economic Perspectives*, Fall 2002.

Knudsen, K. (1980). *Ulikhet i grunnskolen.* Bergen: Universitsforlaget.

Krueger, A. and Lindahl, M. (2002) “The School’s Need for Resources - A Report on the Importance of Small Classes,” The Expert Group on Public Finance (ESO), Stockholm, May 2002.

Krueger, A. B. (2003). Economic Considerations and Class Size, *Economic Journal, Economic Journal, Vol*. 113 (February), pp. F34-F63.

Lie, S. and Turmo, A. (2004). “Hva kjennetenger norske skoler som skårer høyt i PISA 2000?” Acta Didactica no 1 2004, pp. 1-65.

Læreplan L97 ‘Prinsipp og retningslinjer for opplæringa i grunnskulen’, s.81

Læringssenteret. (2004). *Om Statistikken 2003 (Karakterstatistikken 10.klasse)*, http://www.ls.no/Stati/KarG/03/side07.asp

Machin, S. and McNally, S. (2003). “The literacy hour”, mimeo, Centre for Economics of Education, London School of Economics, London.

OECD. (2000). Measuring student knowledge and skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy. Paris: OECD.

Paglin, and Rufolo, . (1990). “Heterogenous Human Capital, Occupational Choice, and Female-Male Earnings Differences”, *Journal of Labor Economics*, Vol 8, no. 1, 123-144.

Raaum, O. (2003) Familiebakgrunn, oppvekstmiljø og utdanningskarrierer, i Statistiske analyser nr 60, Statistisk sentralbyrå.

Raaum, O., Salvanes, K. G. and Sørensen, E. Ø. (2003a). ”The Neighbourhood effect is not what it used to be: Has there been equalisation of opportunity across families and communities in Norway? Discussion paper 36/2001, The Norwegian School of Economics. Revised 2003.

Raaum, O., Salvanes, K. G., and Sörensen, E. Ø. (2003b). “The impact of a primary school reform on educational stratification: A Norwegian study of neighbour and school mate correlation.” *Swedish Economic Policy Review*, Vol. 10.

Solon, G. and Chadwick, L.(2002). "Intergenerational Income Mobility among Daughters," *American Economic Review*, Vol. 92, pp. 335-344.

Statistics Norway. (2001). *Utdanningsstatistikk: Dokumentasjon 2000 av den individbaserte utdanningsstatistikken.* Norges offisielle statistikk C645, Statistisk sentralbyrå.

Todd, P.E. and Wolpin, K. I. (2003). “On the specification and estimation of the production function for cognitive achievement.” *Economic Journal, Vol*. 113 (February), pp. F3-F33.

Shonkoff, J. P. and D.A. Phillips (eds) (2000) From Neurons to Neighbourhoods: The Science of Early Childhood Development. Committee on Integrating the

Science of Early Childhood Development. National Academy Press, Washington, D.C.

Wilson, K. (2002) The effects of school quality on income, *Economics of Education Review* 21 (2002) 579-588.

Woessmann, L. (2004) How equal are educational opportunities? Family background and student achievement in Europe and the US, CESifo, Working Paper No. 1162.

De sist utgitte publikasjonene i serien Rapporter
Recent publications in the series Reports

2003/13 D. Fredriksen, K. Massey Heide, E. Holmøy og N.M. Stølen: Makroøkonomiske virkninger av endringer i pensjonssystemet. 2003. 91s. 180 kr inkl.mva. ISBN 82-537-5173-7

2003/14 B. Aardal, H. Valen, R. Karlsen, Ø. Kleven og T.M. Normann: Valgundersøkelsen 2001. 2003. Dokumentasjon- og tabellrapport. 183s. 260 kr inkl. mva. ISBN 82-537-6408-1

2003/15 A. Finstad, G. Haakonsen og K. Rypdal: Utslipp til luft av partikler i Norge. Dokumentasjon av metode og resultater. 2003 45s. 155 kr inkl. mva. ISBN 82-537-6424-3

2003/16 A. Snellingen Bye, G.I. Gundersen og J.K. Undelstvedt: Resultatkontroll i jordbruk 2003. Jordbruk og miljø. 2003. 95s. 180 kr inkl. mva. ISBN 82-537-6429-4

2003/17 R. Straumann: Exporting Pollution? Calculating the embodied emissions in trade for Norway. 2003. 33s. 155 kr inkl. mva. ISBN 82-537-6487-1

2003/18 O. Vaage: Yrkesliv eller pensjonisttilværelse. Levekår og tidsbruk i aldersgruppen 62-66 år. 2003. 64s. 180 kr inkl. mva. ISBN 82-537-6499-5

2003/19 T. Bye og P.M. Bergh. Utviklingen i energiforbruket i Norge i 2002-2003. 2003. 42s. 155 kr inkl. mva. ISBN 82-537-6508-8

2003/20 B. Halvorsen og R. Nesbakken: Hvilke husholdninger rammes av høye strømpriser? En fordelingsanalyse på mikrodata. 2003. 23s. 155 kr inkl.mva. ISBN 82-537-6511-8

2003/21 T. Bye, P.V. Hansen og F.R. Aune: Utviklingen i energimarkedet i Norden i 2002-2003. 2003. 39s. 155 kr inkl.mva.ISBN 82-537-6513-4

2003/22 Y. Lohne og H.Nome Næsheim: Kartlegging av bruken av deltid i arbeidslivet. 2003. 61s. 180 kr inkl.mva. ISBN 82-537-6521-5

2003/23 A. Snellingen Bye, O. Rognstad og L.J. Rustad: Klassifisering av driftsenhetene i jordbruket etter driftsform og størrelse. 2003. 61s. 180 kr inkl.mva. ISBN 82-537-6529-0

2003/24 R. Nygaard Johnsen: Konsumprisindeks for Svalbard 2003. 2003. 36s. 155 kr inkl.mva. ISBN 82-537-6532-0

2003/25 T.P. Bøe. Funksjonshemmede på arbeidsmarkedet. 2003. 47s. 155 kr inkl. mva. ISBN 82-537-6543-6

2004/1 B. Lie: Ekteskapsmønstre i det flerkulturelle Norge. 2003. 120s. 210 kr inkl.mva.ISBN 82-537-6550-9

2004/2 J. Epland, V. Pedersen, M.I. Kirkeberg og A. Andersen: Økonomi og levekår for ulike grupper, 2003. 2004. 90s. 180 kr inkl. mva. ISBN 82.537-6556-8

2004/3 D. Spilde og K. Aasestad: Energibruk i norsk industri 1991-2001. 2004. 52s. 180 kr inkl. mva. ISBN 82-537-6562-2

2004/4 A. Gillund: Prisindeks for kontor- og forretningseiendommer. 2004. 31s.155 kr inkl. mva. ISBN 82-537-6566-5

2004/5 A. Finstad, K. Flugsrud, G. Haakonsen og K. Aasestad: Vedforbruk, fyringsvaner og svevestøv. Resultater fra Folke- og boligtellingen 2001. Levekårsundersøkelsen 2002 og Undersøkelse om vedforbruk og fyringsvaner i Oslo 2002. 78s. 180 kr inkl. mva. ISBN 82-537-6568-1

2004/6 R.H. Kitterød og R. Kjelstad: Foreldres arbeidstid 1991-2001. 2004. 78s. 180kr inkl. mva. ISBN 82-537-6574-6

2004/7 A. Rolland, Ø. Brekke, B.M. Samuelsen og P.R. Silseth: Evaluering av kommunale brukerundersøkelser. Prosjekt utført for Kommunal- og regionaldepartementet av Statistisk sentralbyrå og Handelshøyskolen BI. 2004. 103s. 210 kr inkl.mva. ISBN 82-537-6582-7

2004/8 T.E. Haug: Eierkonsentrasjon og markedsmakt i det norske kraftmarkedet. 2004. 39s. 155 kr inkl. mva. ISBN 82-537-6597-5

2004/9 M. Kalvarskaia og A. Langørgen: Capital costs in municipal school buildings. 29s. 150 kr inkl.mva. ISBN 82-537-6591-6