Linn Renée Naper

Educational efficiency and institutions

Four essays on the economics of education

Thesis for the degree philosophiae doctor

Trondheim, April 2008

Norwegian University of Science and Technology
Faculty of Social Sciences and Technology
Department of Economics

NTNU

NTNU
Norwegian University of Science and Technology

Thesis for the degree philosophiae doctor

Faculty of Social Sciences and Technology
Department of Economics

ISBN 978-82-471-7846-1 (printed version)
ISBN 978-82-471-7863-8 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2008:93

Printed by NTNU-trykk

To my family

Preface

This thesis consists of an introductory chapter and four essays. The essay in Chapter 2 is joint work with my supervisor, Lars-Erik Borge (Norwegian University of Science and Technology), and the essay in Chapter 4 is joint work with Hans Bonesrønning (Norwegian University of Science and Technology). The essay in Chapter 2, entitled Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools, has been published in *FinanzArchiv*.

Acknowledgements

The prolonged and arduous process of writing this doctoral thesis more than once has challenged me with a feeling of inadequacy and frustration, and there is one particular ingredient in this process that cannot be emphasized and appreciated enough: the people that held my hand along the way. This thesis is a result of the effort from a long list of people. Their involvement and various inputs and suggestions have been of invaluable importance to me. These lines belong to them.

First and foremost, I am deeply indebted to Lars-Erik Borge and Torberg Falch for their guidance and support, confidence and inspiration. Lars-Erik has been my supervisor and he also is my co-author on Chapter 2. Lars-Erik has contributed to this thesis by letting me draw on his vast knowledge within the field of economics and thorough insight in the workings and institutional settings of the Norwegian local public sector. He has this incredible knack of making me see things in their proper perspectives. His feedback is always constructive and well-founded, and he has taught me a lot during the last four years. Torberg has been my co-supervisor. He also guided me through my master thesis and introduced me to the research world. Torberg has a remarkable amount of energy and devotion to his field of research, and he has always found time for me during the last four years. I have benefited a lot from our numerous talks and discussions, and he has been a great inspiration and support for me.

I am also grateful to my colleagues at the Department of Economics at NTNU for insightful comments on seminars and for providing a stimulating and friendly working environment. Hans Bonesrønning is my co-author in Chapter 4 and I have appreciated and benefited a lot from discussing with him. Jørn Rattsø and Bjarne Strøm have contributed with valuable comments on earlier drafts of Chapter 3. I am thankful to Øystein Røkke and Gerd Helene Holm at the Department, and Vigdis Øvrebø at SNF, for providing rapid and effective help with all practical matters. I have appreciated daily coffee breaks, encouragements, good laughs and discussions about important and less important issues with my good friends and fellow students; Marte, Marianne, Silje, Jørgen, Jon, Silje, Nina, Asle, Wenting, Ivar, Jon Olaf, Sofia and Laia.

During fall 2005 and spring 2006 I had the privilege to visit the University of Amsterdam. I am grateful to Professor Hessel Oosterbeek for admitting me in the SCHOLAR group. I learnt a lot from participating in this excellent research environment, and I am grateful to my colleagues at SCHOLAR for their hospitality and friendship. At an early stage I benefited from discussing with Professor Sverre A. C. Kittelsen about the DEA method used in Chapter 2 and 3. Kjell J. Sunnevåg has administrated the survey on organizational approaches in local public service provision used in Chapter 3. I would also like to thank the Ministry of Local Government and Regional Development for funding through a larger project on efficiency in the local public sector, and the members of the project reference group for insightful comments and discussions about the workings of the local public sector.

I will also thank my parents and sisters, Lena, Lise Lotte and Lill Carine, family and friends, for their care and encouragement. Finally, I am grateful to my boyfriend, Ove Andreas, for his affection, ever-lasting patience and support, and for always reminding my of what life really is about.

Linn Renée Naper
Trondheim, December 2007

Contents

Preface
Acknowledgements

Chapter 1:	Introduction and Summary	1
Chapter 2:	Efficiency Potential and Efficiency Variation in Norwegian lower Secondary Schools, with Lars-Erik Borge. *FinanzArchiv 62 (2), pp. 221-249.*	13
Chapter 3:	Teacher Hiring Practices and Educational Efficiency	45
Chapter 4:	Heterogeneous Effects from School Competition, with Hans Bonesrønning	73
Chapter 5:	Gender Gaps in Student Achievement: Evaluation Schemes and Teacher Characteristics	109

Chapter 1
Introduction and Summary

1.1 Introduction

Compared to the average level for the OECD countries, public spending on education in Norway is high, and international student assessment surveys rank Norwegian students only slightly above the average (OECD 2001; 2004)[1]. It is a common opinion that Norwegian students get mediocre results relative to the spending level on educational services (Clemet, 2001; Utdanningsdirektoratet, 2005). Comparable countries, such as other Nordic countries, invest fewer resources per students in compulsory education and are ranked higher in international student assessment rankings (OECD, 2007; Lie et al., 2001; Kjærnsli et al., 2004). These rather discouraging facts represent a major concern and a hot topic in the political and academic debate.

In Norway, as in most countries, the public sector is heavily involved in the provision of schooling. As stressed by Hanushek (2002), this makes understanding the efficiency of production an important issue. If schooling alternatively were provided by private markets, there would generally be less concern about efficiency of production, as market forces in this case are trusted to push the producers towards more efficient use of resources. However, with government involvement in production, many schools operate in a near-monopoly situations and the focus is therefore substantially altered. A key finding in cross-country studies based on international knowledge tests is that that the richest countries, which allocate most resources to the educational sector, do not receive high achievement in return. This result demonstrates the existence of serious inefficiencies in provision, and moreover that there is no relationship between resource use in education and student performance. Studies based on within-country variation also indicate that increased resource use in education is no guarantee for improved results[2]. Empirical studies concerned with the relationship between resource use and student achievement are generally plagued with serious endogeneity problems, and recently a large number of papers have applied convincing empirical identification strategies

[1] Former Minister of Education Kristin Clemet claims in a newspaper article that Norwegian schools are ‘quantitatively best’ but only modest qualitatively, compared to other countries in the OECD, (Clemet, 2001). Official OECD indicators show that Norway spent 9 476 US dollar per student in lower secondary education, as apposed to the OECD average of 6 909 US dollar per student in 2004 (OECD, 2007). Moreover, the average reading test score for Norwegian students at the PISA 2000 was 505 compared to the OECD average of 500. Finland, who came out best of the OECD countries, had an average score of 546.

[2] A review of the literature on the relationship between resource use and achievement can be found in for instance Hanushek (2002) or Wößmann (2005).

to overcome these problems and to pin down the causal effect of increased resource use on educational performance. Despite advanced and well developed methods, the perception among economists, based on prevailing evidence, is still that the direct effect of increased resources in education is zero or small at best.

A suggested explanation to the weak empirical relationship between resource use and educational performance is that policy decisions, such as increased direct spending and reduced class-size, are made within the context of prevailing institutional settings that do not impose strong incentives to increase performance (Hanushek, 2002). Economic research on educational production is mainly focused on sorting out the causal effects of school factors that can be manipulated through policy, from other contributory factors to educational performance[3]. Of current interest is the role played by existing organizational and institutional settings. Incentives are created by rules and regulations in the educational sector, and may exercise an important impact on educational performance and efficiency by imposing rewards for high performance and penalties for low performance (e.g. Hanushek and Wößmann, 2007; Hanushek, 1994). However, existing incentives may also have opposite or mixed effects on performance. Existing empirical evidence suggests that policies such as increased school autonomy and decentralization, competition, high stake testing of students and educational accountability systems may raise strong incentives for increased performance (see for instance Hoxby, 2000; Hoxby and Rockoff, 2004; Hanushek and Raymond, 2005).

The most common conceptual framework employed in economic analyses of education takes the form of a production function, with students' test scores as the typical output and characteristics of the teaching and learning environment as typical inputs (e.g. Todd and Wolpin, 2003). A production function is generally based on an assumption of efficiency, meaning that all institutions should be able to transform inputs to outputs at the same rate. Hanushek (1986) argues that the empirical application of the conceptual model may collapse if inefficiencies exist in the production, which also may explain the failure of identifying any effects of resource use. A related literature, summarized by Worthington (2001), considers the existence of inefficiencies in the educational production process by taking a conceptually different approach using a variety of techniques (econometric or mathematical programming) to identify efficient educational institutions and compare them with inefficient institutions.

[3] Starting with the report by Coleman et al. (1966), 'The Coleman Report', numerous contributions has emphasised the relevance of individual student characteristics, family background, peers etc for individual student performance (e.g. Cook and Evans, 2000; Hanushek et al., 2003; Hægeland et al., 2004).

Although also criticized for certain methodological shortcomings[4], these studies clearly have illustrated the existence of substantial variation in efficiency among different educational institutions.

The aim of this thesis is to contribute to the understanding of the variation in educational efficiency and the impact of existing institutional settings within the education sector. There are few empirical analyses on educational efficiency in Norwegian schools, and the analyses presented in Chapters 2 and 3 contribute to the literature by calculating the efficiency potential of Norwegian schools and school districts (municipalities) using Data Envelopement Analysis (DEA), and by analyzing the determinants of efficiency. Chapter 2 focuses on the impact of economic, demographical and political characteristics of the municipalities, while the analysis in Chapter 3 extends the analysis by including information on the organizational approach taken by the municipalities. The two remaining papers are also based on available data from lower secondary schools in Norway, and apply the traditional production function approach. The topic of Chapter 4 is the impact of competition from private schools on public school outcomes, while Chapter 5 is concerned with the heterogeneity in student achievement measures from different evaluation schemes used in Norwegian education.

1.2 Summary of the Essays

Chapter 2: Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools

This chapter analyses the efficiency potential in the lower-secondary-school sector in Norway. Our aim is to calculate the gain that could be achieved if all municipalities (schools districts) operated their school sector according to the best Norwegian practice. The efficiency potential is identified by comparing performance and resource use among Norwegian schools. The analysis is related to a large literature, starting with Bessent et al. (1982) and summarized by Worthington (2001), which uses DEA to calculate efficiency in the educational sector. We use student achievement in core subjects as outputs, and the number of teacher hours as inputs.

Compared to the international literature, we make two contributions. The first contribution is related to the handling of family background as an indicator of the quality of the students. Exploiting a rich data set of more than 100,000 students containing grades and extensive

[4] The methodological challenges with parametric and non-parametric efficiency frontier techniques are discussed by for example Hjalmarsen et al. (1996).

information on family background, we are able to estimate a measure of student performance adjusted for variation in family background. The adjusted grades may be interpreted as the average grade in the municipality adjusted for family background. We use the adjusted grades, rather than the raw grades, as output variables in the DEA analysis. The advantage of this approach is that the inputs in the DEA analysis can be restricted to factors under direct control of the educational institution and that differences in family background are taken into account in the calculation of the efficiency potential. The second contribution is that we provide an extensive analysis of variation in efficiency scores along the lines of Duncombe et al. (1997) and Grosskopf et al. (2001).

The efficiency potential is calculated to 14 percent. This means that if all municipalities operated their school sector according to the Norwegian best practice, the average resource use could be reduced by 14 percent without reducing student achievement levels. The analysis of the determinants of efficiency indicates that a high level of municipal revenue, a high degree of fragmentation of the political parties, and a high share of socialists in the local council are associated with low educational efficiency. The negative effects of the share of socialists and party fragmentation seem to reflect both higher resource use and lower student performance.

Chapter 3: Teacher Hiring Practices and Educational Efficiency

In recent years, there has been increasing international interest in decentralized decision making within the education sector, and actions have been taken to increase school autonomy in the belief that relieving schools from central regulation will result in better school management and increased efficiency in the educational sector. In a closer examination of the selection process of school teachers, this chapter extends the analysis in Chapter 2 by evaluating whether educational efficiency may be increased through a more effective evaluation and hiring process of teachers. The focus of the analysis is how decentralized hiring affects educational efficiency. Decentralized hiring means that the appointment decision is made at the school level by the principal whereas centralized hiring means that the decision is made by the school district administration. School principals may know better than the school administration which qualifications are most needed for a given teaching position, and may therefore more effectively evaluate the candidates. Decentralized hiring may thus result in an overall better match between schools, teachers and students and is therefore expected to improve educational efficiency.

According to the raw data, efficiency is the highest in districts where hiring is decentralized. There is no direct information about factors that trigger the decision to decentralize, and the causal effect of decentralized hiring may be hard to pin down. If the decision is based on unobservable characteristics, naïve estimates of the effect of hiring on efficiency may be biased because of non-random selection. In the empirical analysis, this problem is first approached by including a large set of controls in a school district level Tobit analysis, which does not alter the qualitative result that school districts that decentralize are more efficient. However, if the decision to decentralize is based on unobservable school district characteristics that are correlated with educational efficiency, the Tobit estimates may still be biased.

Therefore, in a second approach I allow for heterogeneous treatment effects for individual schools in a school level analysis exploiting that the impact of decentralized hiring is expected to be stronger when schools have excess teacher supply. By including school district fixed effects I can estimate the interaction effect between decentralized hiring and teacher supply. The empirical results indicate, as expected, that the effect of decentralization is stronger for schools facing excess teacher supply than for schools without excess supply.

Chapter 4: Heterogeneous effects from school competition

In addition to increasing school autonomy, many countries have introduced reforms to increase the competition from charter schools as a mean to improve the performance of public schools. The education market seems to be a differentiated product market, where schools differ in their characteristics and where parental preferences over school characteristics differ, and it is a controversial issue whether school competition will work as intended. One conjecture is that the degree and character of competition will vary with the characteristics of both suppliers and demanders. In this chapter we evaluate whether the performance of public schools in Norway is affected by competition from private schools, and particularly whether competition has heterogeneous effects across student subgroups.

The existing empirical literature on school competition- between public and private schools and between public schools – does not provide any sweeping and clear-cut conclusions (e.g. Hoxby, 2000; Fiske and Ladd, 2001;Dee 1998). This chapter provides empirical evidence from Norway. First, we investigate whether the overall performance of public schools is affected by private school competition. Second, we ask whether the competition effects differ across student subgroups. If public sector actors’ response to competition is to reallocate resources between student subgroups, we expect that attractive and mobile student subgroups gain from competition, while less attractive and less mobile students lose from competition with private schools. The endogeneity of competition - measured by the share of students attending private schools – is treated by IV estimation. Sorting of parents across municipalities is considered by exploiting the location of universities and colleges. Indications of non-uniform competition effects are provided: while there is no evidence that students with less educated parents gain from competition, the effect for students with highly educated parents is positive. The robustness analysis indicates that the effect to some extent depends on the outcome variable used.

Chapter 5: Gender gaps in student achievement: evaluation schemes and teacher characteristics

Such as most empirical analyses of the education market, the analyses presented in this thesis exploit student achievement in terms of marks and test scores to measure educational production and efficiency. Educational efficiency in Chapter 2 and 3 is measured using individual student achievement based on teacher assessment, while the analysis of school competition in Chapter 4 also exploits individual student achievement from an anonymously evaluated national one-day test. Different testing schemes may have heterogeneous effects across student subgroups, and this chapter assesses this heterogeneity by comparing individual student achievement across gender from three different evaluation schemes.

Student achievement in terms of marks and test scores is frequently used as an indicator of educational production and performance. Individual outcomes from different evaluation schemes are also important in application for higher education as well as for future job prospects. It is therefore important to ensure that student achievement is evaluated and measured as correctly as possible. This chapter evaluates individual student achievement levels from three different evaluation schemes in Norway. First, using difference in difference estimation, the analysis reveals substantial gender biases in teacher assessment. Conditional on individual test scores from anonymously evaluated central exams, girls receive significantly higher marks than boys when assessed by the teacher. Exploiting intertemporal variation within schools, teachers are found to assess same-sex students less favourably than opposite-sex students in Norwegian. Hence, the observed gender bias is related to teacher grading practices. Second, comparing individual student performance from a high stake central exam and a low stake national test, girls are found to perform significantly better when stakes are high. Girls' relative ability to perform when exposed to a high stake one-day test is therefore not a likely explanation for the observed gender bias in teacher assessment.

1.3 Discussion and Future Research Agendas

This thesis consists of four empirical analyses dealing with different topics within the field of economics of education, and contributes to the literature by improving on the measurement of educational efficiency by means of Data Envelopement Analysis, and by raising important and new questions concerning the organization of the school sector. The education market is characterized by large heterogeneity among both suppliers and demanders, making the educational production process complicated. The heterogeneity is clearly reflected by the key findings of this thesis. First, efficiency is shown to vary substantially both across school districts, and across individual schools. Second, the analyses find that efficiency depends on teacher hiring practices in the school districts as well as the degree of competition from private schools, but the effects are heterogeneous across schools and students respectively. Moreover, existing student achievement evaluation schemes are found to produce heterogeneous outcomes depending on the gender of the student.

Numerous challenges and unsolved questions remain in the literature, and the thesis is by no means meant to be exhaustive in illuminating all aspects of the production process. Based on today's empirical evidences the key for future improvements and increased efficiency in educational production seems to lie in better incentives, to promote better management and to staff schools with high quality teachers. Still, the literature provides relatively few guidelines on how to develop good educational institutions with funding schemes and incentive systems that may ensure efficient use of resources and improve student achievement levels. More research on this topic is therefore highly demanded, and the objective for future research should be to expand our understanding of the interplay between resource use and the organizational and institutional approaches taken within the educational sector.

References

Bessent, A., Bessent, W., Kennington, J. and Reagan, B. (1982), An Application of Mathematical Programming to Assess Productivity in the Houston Independent School Districts, Management Science 28, 1355-1367

Clemet, K (2001): Arbeids- og administrasjonsdepartementet, Newspaper Article, Verdens Gang 6.12.2001.

Coleman, J. S., Campbell, E. Q, Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. and York, R. L. (1966): Equality of Educational Opportunity: Washington, D.C.:U.S. Government Printing Office.

Cook, M. D. and Evans, W. N. (2000): Families or schools? Explaining the convergence in white and black academic performance, Journal of Labour Economics, 18 (4).

Dee, T. S., (1998): Competition and the Quality of Public Schools, Economics of Education Review, 17, pp. 419-427

Duncombe, W., Miner, J. and Ruggiero, J. (1997), Empirical Evaluation of Bureaucratic Models of Inefficiency, Public Choice 93, 1-18.

Fiske, E. B. and Ladd, H., (2001): Does Competition Generate Better Schools? Evidence from New Zealand, Working Paper Series SAN01-16, Terry Sanford Institute of Public Policy.

Grosskopf, S., Hayes, K.J., Taylor, L.L. and Weber, W.L. (2001), On the Determinants of School District Efficiency: Competition and Monitoring, Journal of Urban Economics 49, 453-478.

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

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

Hanushek, E. A (2002): Publicly Provided Education, National Bureau of Economic Research, Working Paper 8799.

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

Hanushek, E. A. and Raymond M. E. (2005): Does school Accountability Lead to Improved Student Performance?, Journal of Policy Analysis and Management, 24(2), pp. 297-327.

Hanushek, E. A. and Wößmann, L. (2007): The Role of School Improvement in Economic Development, National Bureau of Economic Research, Working Paper 12832.

Hjalmarsson, L., Kumbhakar, S.C. and Heshmati, A. (1996): DEA, DFA and SFA: A comparison, Journal of Productivity Analysis 7, pp. 303-327.

Hoxby, C.M. (2000), Does Competition Among Public Schools Benefit Students and Taxpayers?, American Economic Review 90, 1209-1238.

Hoxby, C. M., and Rockoff J. E., (2004): The impact of Charter Schools on Students Achievement, Working paper Department of economics, Harvard University, Cambridge.

Hægeland, T., Raaum, O. and Salvanes, K.G. (2004): Pupil Achievement, School Resources and Family Background, Mimeo, Ragnar Frisch Centre for Economic Research and Statistics Norway.

Kjærnsli, M., Lie, S., Olsen R.V., Roe, A. og Turmo, A. (2004): Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003,Universitetsforlaget.

Lie, S., Linakulä, P. and Roe, A. (2001): Northern Lights on PISA: Unity and diversity in the Nordic countries in PISA 2000. Department of teacher education and school development, University of Oslo, Norway.

OECD (2001): Knowledge and Skills for Life. Paris: OECD Publications.

OECD (2004): Learning for Tomorrow's World – First Results from PISA 2003 Paris: Organization for Economic Co-operation and Development.

OECD (2007): Education at a Glance, OECD Indicators 2007, Organization for Economic Co-operation and Development.

Todd, P. E. and Wolpin, K. I. (2003): On the Specification and Estimation of the Production function for Cognitive Achievement, The Economic Journal, 113, F3-F33.

Utdanningsdirektoratet (2005): Utdanningsspeilet 2004: Analyse av grunnskole og videregående opplæring i Norge.

Worthington, A. C. (2001): An empirical Survey of Frontier Efficiency Measurement Techniques in Education, Education Economics, 9 (3).

Wößmann, L. (2005): Educational Production in Europe, Economic Policy, 20 (43), pp. 446-504.

Chapter 2
Efficiency Potential and Efficiency Variation in Norwegian lower Secondary Schools

Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools

Lars-Erik Borge and Linn Renée Naper*

Received 15 November 2005; in revised form 23 March 2006; accepted 4 April 2006

The paper performs an efficiency analysis of lower secondary schools in Norway. The efficiency potential is calculated as 14%, based on a DEA analysis with grades in core subjects (adjusted for student characteristics and family background) as outputs. The analysis of the determinants of efficiency indicates that a high level of municipal revenue, a high degree of party fragmentation, and a high share of socialists in the local council are associated with low educational efficiency. The negative effects of the share of socialists and party fragmentation seem to reflect both higher resource use and lower student performance.

Keywords: educational efficiency, DEA analysis, determinants of efficiency, political and budgetary institutions

JEL classification: I 21

1. Introduction

The educational sector has received substantial attention in academic and political debate in recent years. International knowledge tests have provided new and easily accessible information that facilitates a comparison of educational performance across countries. A key finding is that the international tests indicate a negative correlation between student performance and resource use, meaning that the richest countries, which allocate the most resources to the educational sector, do not receive high achievement in return. Norway, which is a rich country with high resource use and (at

* Earlier versions of this paper have been presented at the Congress of the European Economic Association (Amsterdam), the Norwegian–German Seminar on Public Economics (Garmisch-Partenkirchen), the Meeting of the European Public Choice Society (Durham), the Meeting of the Norwegian Economic Association (Oslo), and seminars at the Norwegian University of Science and Technology (Trondheim) and at the Institute for Research in Economics and Business Administration (Bergen). Comments and suggestions from the participants are gratefully acknowledged, in particular from Torbjørn Hægeland, Alfons Weichenrieder, Arnaud Bilek, and Bjarne Strøm; and also from two referees. We have also benefited from discussions with Sverre A. C. Kittelsen about the DEA method. The research is funded by the Ministry of Local Government and Regional Development, and is part of a larger project on efficiency in the local public sector.

FinanzArchiv 62 (2006), 221–249 doi: 10.1628/001522106X120677
ISSN 0015-2218

best) average student performance, is no exception to this tendency. The mismatch between resource use and performance has triggered a political debate regarding resource use, curriculum, and the organization of the school sector. Measures have been taken to increase the number of hours in basic subjects, to provide better information on the performance of individual schools, and to open public schools to more competition from private schools.

The purpose of this paper is to analyze the efficiency potential in the lower-secondary-school sector in Norway. The efficiency potential is not identified by comparing Norway with other countries, but rather by comparing performance and resource use among Norwegian schools. Our aim is to calculate the gain that could be achieved if all municipalities operated their school sector according to the best Norwegian practice. The analysis is related to a large literature, starting with Bessant et al. (1982) and summarized by Worthington (2001), which applies data envelopment analysis (DEA) to the educational sector. To our knowledge this is the first DEA analysis of lower secondary education in Norway that uses grades or student achievement as outputs.[1]

With respect to the existing international literature, we make two contributions. The first relates to the handling of family background as an indicator of the quality of the students. We take advantage of a rich data set of more than 100,000 students containing grades and extensive information on family background to estimate a measure of student performance adjusted for variation in family background. It is these adjusted grades, rather than the raw grades, that are used as outputs in the DEA analysis. The advantages of this approach are that the inputs in the DEA analysis can be restricted to factors under direct control of the educational institution and that differences in family background are taken into account in the calculation of the efficiency potential. The second contribution is that we provide an extensive analysis of variation in efficiency scores along the lines of Duncombe, Miner, and Ruggiero (1997) and Grosskopf et al. (2001). We investigate the influence of political and budgetary institutions of the municipality, along with traditional variables like school size.

In Norway primary and lower secondary education is mainly a municipal responsibility, and the municipalities are the units of observation in this study. The DEA analysis reveals large variations in efficiency across municipalities, and the average efficiency potential is calculated as 14%. The variation in efficiency is analyzed using tobit regression, which indicates that a high level of revenue, a high degree of party fragmentation, and a high share of socialists in the local council are associated with low educational efficiency.

1 Bonesrønning and Rattsø (1994) analyze Norwegian high schools using grades as output.

The negative effects of the share of socialists and party fragmentation seem to reflect both higher resource use and lower student performance.

The rest of the paper is organized as follows: Section 2 provides the necessary institutional background. The principles of DEA analysis and the approach taken in this paper are discussed in section 3; section 4 discusses data and model specification. Section 5 presents the findings of the DEA analysis and discusses the robustness of the results. Section 6 is devoted to the tobit analysis of the determinants of educational efficiency. Finally, section 7 offers some concluding remarks.

2. Institutional Background

Most primary and lower secondary schools in Norway are owned and operated by the municipalities. Private schools account for less than 2% of the students, and until the school year 2003/04 the few private schools in operation were either religious schools or schools that use alternative educational methods. In 2003 the parliament passed a new law on private schools, which allows for nonreligious private schools that use traditional educational methods. This study is based on data for the school years 2001/02 and 2002/03 and only includes municipal schools.

Norwegian municipalities are multipurpose authorities that, in addition to education, are responsible for welfare services like child care, primary health care, and care for the elderly. Other important tasks are culture and infrastructure. The main revenue sources are taxes (43% of current revenue), block grants (21%), earmarked grants (13%), and user charges (16%). Interest and other revenue account for the rest. Since earmarked grants and user charges are practically nonexistent in primary and lower secondary education, the sector is mainly financed by taxes and block grants. Compared to most other countries, the system of financing is quite centralized. Around 95% of local taxes are regulated income and wealth taxes where effective limits on tax rates have been in place for the last 25 years. The opportunity to influence current revenues is limited to property tax and user charges.

The municipalities enjoy more discretion on the spending side than on the revenue side. The allocation of taxes and block grants between different service sectors is decided locally, subject to national regulations and minimum standards. In the educational sector there is a national curriculum, and minimum standards are determined by a maximum class size and minimum number of hours per class in each subject.[2] Moreover, until 2004 the teachers' unions negotiated wage and workload (teaching hours per week) with

2 From the school year 2003/04 the class size regulation was replaced by a more flexible regulation of group size.

the national government. Despite these national regulations, there is substantial variation in resource use per student between schools and between municipalities (see section 4). The variation between schools is to a large extent related to school size, and the variation between municipalities is related to the choice of school structure and thereby average school size. The settlement pattern, the number of students in the municipality, and the municipal revenue are important determinants of average school size. Few students and a decentralized settlement pattern tend to give small schools, and municipalities with high levels of revenue can afford a decentralized school structure with many small schools.

3. Data Envelopment Analysis

Over the last decades several methods have been developed to estimate production frontiers and the degree of efficiency for each unit of production. Today the two dominant approaches are data envelopment analysis (DEA) and stochastic frontier analysis (SFA).[3] The two approaches have different strengths and weaknesses. DEA is a linear programming method that has the advantage that no restrictive assumptions about technology (the functional form of the production function) or the distribution of efficiency have to be made, and that it easily handles multiple inputs and outputs. The main weakness of DEA stems from the fact that it is a nonstatistical method with no random error. As a consequence, it does not produce statistical tests and is sensitive to measurement error. SFA is an econometric approach that allows for statistical testing and is less sensitive to measurement error, but requires the researcher to impose a specific functional form of the production function and make strong assumptions about the distributions of the random error and the efficiency term.

There is no consensus in the literature on the choice between DEA and SFA, and both approaches are widely used in applied work. Hjalmarsson, Kumbhakar, and Heshmati (1996, p. 304) conclude that "the choice between different approaches must be based on trade-offs concerning the purpose of the study, type of data, technology characteristics, etc." In this paper we rely on DEA. The DEA method is attractive in our case because the knowledge of the functional form of the educational production function is limited and because the data set allows for multiple inputs and outputs. Moreover, our primary interest is to calculate and explain variations in educational efficiency, not to provide estimates of the educational production function.

3 De Borger and Kerstens (1996) and Hjalmarsson, Kumbhakar, and Heshmati (1996) provide comparisons of DEA and SFA.

The DEA approach was first introduced by Charnes, Cooper, and Rhodes (1978) in order to calculate relative technical efficiency in the case of multiple inputs or outputs in the production process of nonprofit actors, e.g., in the public sector. Technical efficiency is a normative concept and should be interpreted as the inputs or outputs compared to a standard or a norm, and the basic concept of the DEA procedure used in this paper is to minimize the level of inputs for a given amount of outputs.[4] This is done by simultaneously solving a linear programming problem for each unit (municipalities in our case). Generally municipalities will value their inputs and outputs differently and thus call for different sets of weights in the conventional measure of relative efficiency.[5] The efficiency of a single municipality is calculated relative to a *best-practice* reference frontier. This frontier is defined as a linear combination of the inputs and outputs of efficient municipalities in the sample. The weights and the efficiency measure for each municipality are identified simultaneously in the DEA procedure. As discussed above, the method requires no *a priori* specification of the functional form of the educational production function.

We will now illustrate the DEA procedure by considering a simplified educational sector where a single input (number of teachers per student) is used to produce a single output (student achievement). The simplification allows us to describe the production process in a simple two-dimensional diagram as in figure 1. The points A, B, C, and K represent locations of different municipalities in the input–output space.

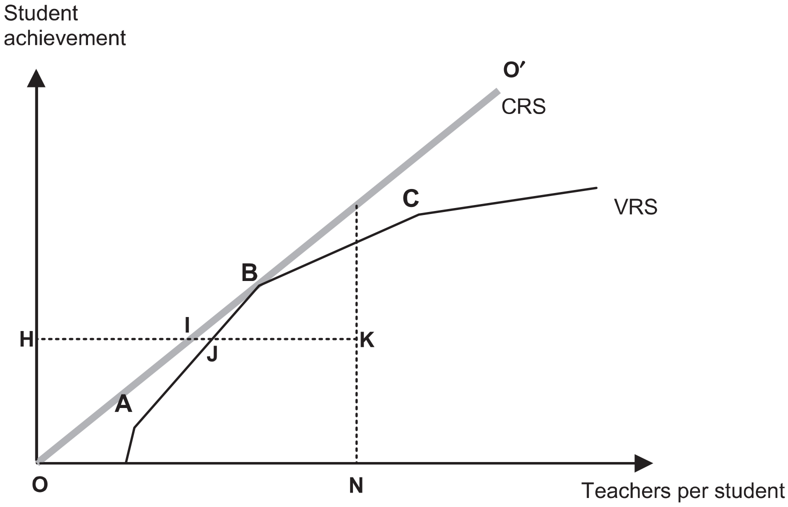
The DEA model originally proposed by Charnes, Cooper, and Rhodes (1978) was input-oriented and allowed only for constant returns to scale. This approach has since been widely developed; see, e.g., Banker, Charnes, and Cooper (1984), where a variable-return-to-scale specification was first proposed. We start out by considering the case with constant returns to scale. In figure 1 the efficiency frontier is then represented by the line OO' passing through the origin and observation B in the diagram.[6] Observation B lies on the reference frontier and is assumed to be fully efficient, while observations that lie below the line OO' (e.g., A, C, and K) are inefficient. Inefficiency implies that the observed units could have produced the same level of outputs with less input by using the best-practice technology defined by the reference frontier. A municipality is less inefficient the larger the

4 Alternatively the efficiency may be calculated by maximizing the outputs for a given level of inputs. In this paper we focus only on input-oriented technical efficiency.

5 Conventionally, relative efficiency is calculated as the ratio of weighted outputs to weighted inputs with a common set of weights.

6 On drawing a line from the origin through each sample observations in figure 1, the line passing through observation B has the greatest slope. Observation B is the most productive of the sample observations. All the other observations lie below this line.

Figure 1
The Best-Practice Reference Frontier under Constant and Variable Returns to Scale



distance to the frontier. In figure 1 the efficiency of observation K can be expressed as the ratio of efficient use of inputs to the actual use of inputs; this ratio is represented in the figure as the distance HI divided by the distance HK. For all observations situated below the efficiency frontier, this ratio will lie between zero and one, while for observation B (the efficient municipality) it is equal to one.

With variable returns to scale the reference frontier is represented by the piecewise linear curve passing through observation A, B, and C in figure 1. In this case only observation K is situated below the efficiency frontier and defined as ineffective. Given the output of municipality K, the efficient amount of inputs is represented by point J, and the relative efficiency (or efficiency score) is thus given by the ratio HJ/HK.

One characteristic of the DEA method is that the number of efficient units and the calculated efficiency potential depend on the number of inputs and outputs relative to the sample size. For a given sample size an increase in the number of inputs and/or outputs will increase the number of efficient units and reduce the calculated efficiency potential. It becomes important to formulate a proper model specification, since an overspecified model (with many outputs and inputs) may underestimate the efficiency potential, whereas an underspecified model (with few outputs and inputs) may overestimate it.

When the DEA method is applied to the educational sector, it is a challenge to limit the number of variables. It is well documented that socioeconomic variables capturing family background are important determinants of student achievement (e.g., Hanushek, 1986), and the potential number of relevant variables describing the quality of student input is very large. In applications of DEA to the educational sector this problem is dealt with in two different ways; see Coelli, Prasada Rao, and Battese (1998) and Worthington (2001). The first is a two-stage procedure where only factors under direct control of the educational institution are included as inputs in a first-stage DEA analysis, and where variables capturing family background are included in a second-stage tobit regression. The problem with this approach is that the efficiency scores from the DEA analysis are biased because differences in family background are not taken into account.

The second alternative is to include variables capturing family background as inputs in the DEA analysis to get unbiased efficiency scores. However, if it is necessary to include a large number of socioeconomic input variables, the efficiency scores may be biased because the number of inputs and outputs becomes large relative to the sample size. The practical solution is to include a few variables or to construct an index of the socioeconomic environment (e.g., Duncombe, Miner, and Ruggiero, 1997). In this case the remaining question is whether family background is sufficiently controlled for.

In this paper we propose a third alternative, where we utilize a rich data set of more than 100,000 10th-graders containing information on grades, student characteristics, and family background. We estimate regressions with individual grades as dependent variable and variables capturing family background as explanatory variables. In addition, a full set of dummy variables for each municipality is included. The estimates of the municipal dummy variables, which may be interpreted as grades adjusted for family background, are used as outputs in the DEA analysis. The advantages of this approach are that the inputs in the DEA analysis can be restricted to factors under direct control of the educational institution and that a large set of variables describing family background can be taken into account in the calculation of the efficiency potential. Moreover, it is not necessary to decide *ex ante* whether each socioeconomic variable has a positive or a negative effect on achievement, as is necessary when the variables are included as inputs in the DEA analysis. A similar approach is used by Grosskopf et al. (2001), where output is based on value-added residuals from a regression with current test scores as dependent variable and previous test scores and the socioeconomic composition of the student body as explanatory variables. However, they do not use data for individual students and only control for a few socioeconomic characteristics.

Another potential problem is that the DEA method is sensitive to measurement error and outliers that tend to overestimate the efficiency potential. The reason is that outliers with high output and/or low input use will affect the position of the frontier and thereby reduce the efficiency score of other units. Outliers with low output and/or high input will only have a minor effect, since they only affect the average efficiency by making themselves less efficient.

In the empirical analysis we use adjusted grades as outputs. In small schools and small municipalities in particular, average grades may vary from year to year, reflecting (unobservable) variation in the quality of the student population. As a consequence the efficiency potential may be overestimated because the frontier is determined by the units with high-quality students. In the empirical analysis we try to reduce this problem by using data that is averaged over two school years. Controlling for student characteristics and family background, as discussed above, also contributes to reducing the problem of variation in student quality. In addition we perform jackknifing to investigate whether the results are sensitive to outliers and measurement error.

4. Data and Specification of the Educational Production Function

Most applications of DEA to the educational sector use grades or test scores as outputs, and in this study we follow that tradition. We take advantage of a national database with information on grades for individual students for the school years 2001/02 and 2002/03. The database provides information on individual assessment grades and exam result for students in the final year (10th grade) of lower secondary education. Written national exams are arranged in the three core subjects Norwegian, English, and mathematics, but each student is only examined in one subject. Even at the municipal level there are many cases where not all three core subjects are covered by national exams. Assessment grades have a wider coverage and are reported for 13 different subjects for each student.

Since the purpose of this study is to investigate variations in efficiency across municipalities, educational output is mainly measured by assessment grades, which are available for all municipalities. The assessment grades (in the final year) are high-stakes tests in that they are used for applications to the upper secondary level. Students therefore have incentives to put effort into home and class work. A possible weakness of assessment grades is that grading practice may vary between municipalities. We think this problem is smaller for the core subjects (Norwegian, English, and mathematics), where national exams are arranged, since the results on national exams represent important feedback to teachers that facilitate correction of divergent grading practices. In the benchmark model we therefore concentrate on assessment grades in the core subjects as indicators of educational output. Some descriptive statistics for the mean assessment grades are given in appendix table 7. Grades are given on a 1–6 scale where 6 is the best. In Norwegian the grade varies from 2.9 in the municipality with the lowest grade to 4.8 in the municipality with the highest grade, with a mean of 3.8. The mean grade is somewhat lower in mathematics than in Norwegian and English, but the variation across municipalities is of roughly the same magnitude in the three subjects.

Although we think assessment grades in core subjects are reasonably good indicators of educational output, we also formulate production functions with additional outputs. First, a possible problem with focusing on core subjects is that we may underestimate the degree of efficiency in municipalities where schools have devoted a large amount of effort and resources to other subjects. To take account of this possibility we formulate a model where the average grades in other subjects are included as an additional output. Second, variation in grading practices is a potential problem even when we focus on core subjects. To take this into account we also formulate a third model where the results of the written national exam are included as an extra output.

Table 7 also reports descriptive statistics for the average assessment grade in other subjects and results from the written national exam. It appears that the average grade level is higher in other subjects than in core subjects, and that the variation across municipalities is smaller. For the exam results the difference goes in the opposite direction: The average grade level on the written exam is lower than the assessment grades in core subjects, and the variation across municipalities is larger.

As discussed in section 3, we utilize a rich data set of more than 100,000 10th-graders to construct grades that are adjusted for family background. The data set contains individual assessment grades and information on socio-economic background for all 10th-graders in the academic years 2001/02 and 2002/03. This rich data set enables us to adjust the grades for family background, and the adjustment is done by performing a regression of the following type:

$$y_{ijt} = \beta x_{ijt} + \gamma_t + \alpha_j + u_{ijt}\,, \tag{1}$$

where y_{ijt} is the assessment grade of student i in municipality j in school year t. The vector x captures student characteristics and family background, γ_t is a year-specific constant term, α_j represents municipal fixed effects, and u_{ijt} is an error term. The α's may be interpreted as the average grade in the municipality adjusted for family background.

Equation (1) is estimated for each of the three core subjects Norwegian, English, and mathematics, for the average grade of the remaining subjects, and for the results of the written national exam. The vector x contains variables that are typically used in analyses of student achievement. It includes a number of individual dummy variables on the student's gender, quarter of birth (given that they graduated in the year they turned 16), graduation earlier or later than expected, and whether they are immigrants or adopted. Family background is captured by parents' education and income (separate for the mother and the father) and dummy variables reflecting whether the parents are married to each other, cohabitants, separated, divorced, or none of these. We do not have information on whether individual students receive adapted teaching due to learning disabilities, only on the fraction of students at each school that receive such teaching. This fraction, labeled the fraction of students with special needs, is also included in x.

It should be noted that the estimated equation for the results from the national exam is slightly different from the estimated equations for the assessment grades. First, because the average grade level on the exam differs between subjects, it is necessary to include controls for whether the student has been examined in Norwegian, English, or mathematics. Second, we have tested for possible peer-group effects by including a variable describing the average level of education for the parents of the other students in the school. It turned out that such peer-group effects were present in the results from the national exam, but not in the equations for the assessment grades.

The appendix table 8 reports descriptive statistics for the x variables; the regression results are reported in table 9. The estimated coefficients in table 9 mainly serve to control student achievement for the available variables capturing family background, and to a great extent they confirm the findings from other analyses on the effect of family background on student achievement; see Hægeland, Raaum, and Salvanes (2004) for a recent Norwegian analysis, and Hanushek (2002) for a survey of international contributions. In brief, we find that parental educational level and income have positive and highly significant effects on student achievement, while immigrants have significantly lower achievement levels. Students living with both parents (either married or cohabiting) get higher grades than students living with only one of their parents (single, separated, or divorced). The estimated effect of the share of students receiving adapted teaching at the student's school is significantly negative and indicates that schools with a higher share of students with special needs have lower achievement level.

The average values of the estimated municipal fixed effects (the α's) are close to zero, and are not directly comparable to the original grades, which are on a 1–6 scale. They are made comparable by adding 3.5 to the α's.[7]

7 The municipal fixed effects also need to be transformed (to be greater than zero) in order to be used as outputs in the DEA analysis. By adding 3.5 this requirement is fulfilled.

Table 1
Descriptive Statistics for Adjusted Grades, Teacher Hours per Student, and the Fraction of Certified Teachers

	Mean	Coefficient of variation	Min	Max
Adjusted grades				
Norwegian	3.55	0.052	2.94	4.34
English	3.53	0.053	3.04	4.19
Mathematics	3.54	0.055	2.99	4.42
Other subjects	3.53	0.042	3.13	4.14
National exam	3.51	0.061	2.73	4.36
Teacher hours per student	96.2	0.279	61.3	226.2
Fraction of certified teachers	0.95	0.057	0.69	1.00

Notes: The figures are based on data for 426 municipalities. The reported means are unweighted averages.

These adjusted grades are reported in table 1. It appears that the variation in adjusted grades is slightly less than the variation in the original grades, which indicates that the adjustment has the expected effect: Municipalities with low grades and poor socioeconomic status are lifted up, whereas municipalities with high grades and good socioeconomic status are leveled down. We also observe that the differences in the coefficient of variation across subjects are reduced compared to the raw grades reported in table 7.

The input measures we use are based on the total number of teacher hours and the fraction of certified teachers (meaning that they have certified education for the relevant grade level). Table 1 documents a substantial variation in teacher hours per student across municipalities, from a low of 61 to a high of 226. On average only 5% of the teachers are noncertified, but in some municipalities up to about 30% of the teaching staff is noncertified.

Table 2 displays the correlations between grades and teacher hours per pupil and the fraction of certified teachers. It appears that the five output measures are positively correlated. The correlations between adjusted assessment grades are in the range 0.4–0.7. The correlations between the assessment grades and the results of the written exam are weaker. However, since the exam result is an aggregate of grades in the three core subjects, the correlation between the exam result and a single assessment grade may be misleading. It is more appropriate to consider the correlation between the exam result and the average assessment grade in the core subjects. This

Table 2
Correlation Matrix for Adjusted Grades, Teacher Hours per Student, and Share of Certified Teachers

	Norwegian	English	Mathematics	Other subjects	National exam	Teacher hours per student	Share of certified teachers
Norwegian	1.000						
English	0.603	1.000					
Mathematics	0.507	0.421	1.000				
Other subjects	0.688	0.619	0.610	1.000			
National exam	0.235	0.312	0.403	0.329	1.000		
Teacher hours per student	0.311	0.319	0.287	0.373	0.213	1.000	
Share of certified teachers	−0.079	−0.089	0.006	−0.078	0.085	−0.241	1.000

correlation is about 0.4, which is of the same order as the correlation between the assessment grades in English and mathematics. The fact that the correlations between the five outputs are clearly less than unity can be also be considered as support for our choice of using the DEA approach rather estimating an SFA model with a single output.

Teacher hours per student are positively correlated with adjusted assessment grades, whereas the fraction of certified teachers is only weakly correlated with the five outputs. The positive correlation between adjusted grades and teacher hours per student is consistent with the results of Hægeland, Raaum, and Salvanes (2004). They find a positive (but modest) effect of teacher hours per pupil on assessment grades after family background is controlled for.

As mentioned above, we use three specifications of the educational production function in the DEA analysis. They have the same specification of inputs, but differ in the specification of outputs. As student characteristics and family background are controlled for in the calculation of adjusted assessment grades, only factors under municipal control are included as inputs. The two inputs that are included are the numbers of teacher hours given by certified and noncertified teachers, respectively. In the benchmark model (A), adjusted grades in the core subjects Norwegian, English, and mathematics are used as outputs. In model B the average grade in other subjects is included as an additional output. Model B has the advantage that it has a more comprehensive output measure, but the disadvantage that grades in other subjects may be less comparable across municipalities due to varying grading practices. The third model (C) extends the benchmark model by including the results of the national exam as an additional output. The advantage of model C is that it includes an output measure (exam results) that is not subject to varying grading practice across municipalities. On the other hand, it has the disadvantage that the results of the national exam are based on results from different subjects.

In the DEA analysis the outputs are specified as the adjusted grades multiplied by the number of students, and the inputs as the total number of teacher hours by certified and noncertified teachers, respectively.[8,9] In all three specifications we allow for variable returns to scale (VRS). Moreover, we focus on input-oriented efficiency scores, because the number of students, which is an important element of the output measures, is largely beyond municipal control. The analysis is based on data for 426 (out of 434) municipalities.

5. Educational Efficiency: The Results of the DEA Analysis

Descriptive statistics for the efficiency scores from models A, B, and C are reported in table 3. In model A, where adjusted assessment grades in core subjects are used as outputs, the mean efficiency score is 0.78 when all municipalities are given equal weight. This means that the average municipality could reduce inputs by 22% without reducing measured output. The results are similar to the U.S. studies by Duncombe, Miner, and Ruggiero (1997) and Grosskopf et al. (2001). Duncombe, Miner, and Ruggiero calculate an average efficiency score of 0.76 in their study of New York State school districts, whereas Grosskopf et al. (analyzing public schools in Texas) find that inputs could be reduced by roughly 20% without reducing output. Kirjavainen and Loikkanen (1998) find average efficiency of 0.82–0.84 in an analysis of senior secondary schools in Finland.

Table 3

Descriptive Statistics for the Calculated Efficiency Scores

	No. of effective units	Mean (unweighted)	Mean (weighted)	Minimum	1st quartile	3rd quartile
Model A	19	0.784	0.860	0.424	0.707	0.873
Model B	20	0.787	0.864	0.424	0.708	0.878
Model C	21	0.793	0.869	0.425	0.716	0.879

8 As a consequence the formulation of the DEA models is slightly different from the illustration in figure 1.

9 The data on the teacher hours only distinguish between primary schools (1st to 7th grade) and lower secondary schools (8th to 10th grade). In the analysis the numbers of teacher hours and students refer to the lower secondary level.

It is the weighted average of the efficiency score (with the number of students as weights) that reflects the national efficiency potential. The weighted average is 0.86, which yields an efficiency potential of 14%. The calculated efficiency potential reflects substantial variation in efficiency score across municipalities. 19 out of 426 municipalities come out as efficient (with an efficiency score of 1), whereas the lowest efficiency score is 0.42. Around 25% of the municipalities come out with an efficiency score below 0.71, and another 25% have an efficiency score above 0.87.

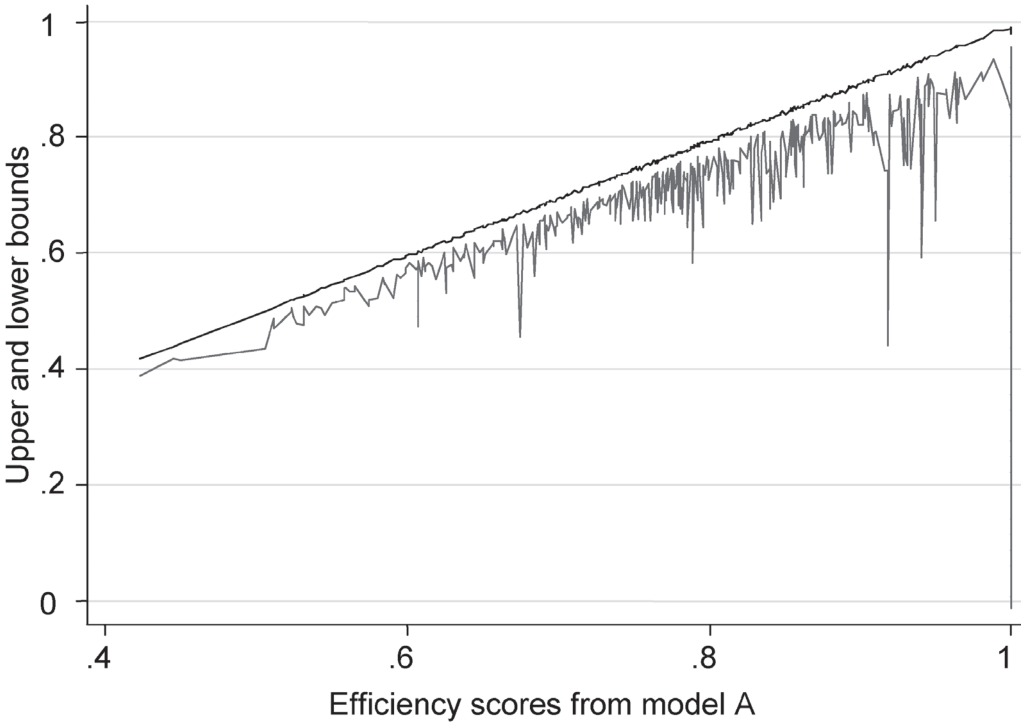
Model B, which includes average grades in other subjects as an additional output, gives more or less the same results as model A. The calculated efficiency potential is slightly reduced, from 14% to 13.6%. The ranking of the municipalities is also largely unaltered: the rank correlation is as high as 0.997. The robustness of the results indicates that high achievement in core subjects does not come at the expense of the achievement in other subjects, and that we do not lose much by focusing on the core subjects.

The results from model C, which includes the result from the written exam as an additional output, are also very similar to the results from model A. The difference in mean efficiency score is less than 1 percentage point, and the rank correlation is as high as 0.990. It might be argued that the high correlation reflects that the assessment grades are the dominating outputs in both models. We have investigated this possibility by comparing model A with a model where the exam result is the only input. Even in this case, where the two models have no common output, the rank correlation is as high as 0.92.

Although the reported quartiles (table 3) offer information about the variation in efficiency across municipalities, the DEA procedure does not provide us with any indication whether the calculated efficiency scores are significantly different between the municipalities. In recent years bootstrapped methods have been developed to calculate standard errors and confidence intervals for the efficiency scores; see Simar and Wilson (1998, 2000, 2004). We perform a smoothed bootstrap,[10] and the confidence intervals of the empirical distribution are obtained by drawing 2000 bootstrap samples. Upper and lower bounds for 95% confidence intervals for each municipality are shown in figure 2, where the municipalities are sorted according to the efficiency scores from model A. The midpoints of the confidence intervals are so-called bias-corrected efficiency scores that are outputs from the bootstrap procedure, where the bias reflects that the municipalities are now compared with a true frontier that lies above the usual DEA frontier. With a few exceptions the confidence intervals are narrow compared to the variation

10 The bandwidth is set according to the rule of thumb proposed by Simar and Wilson (2004).

Figure 2
Upper and Lower Bounds for 95% Confidence Intervals for the Bias-Corrected Efficiency Scores (Model A)



in the efficiency scores, and there are few overlaps between the intervals to the left of the figure and those to the right. Figure 2 can therefore be taken as evidence that the efficiency scores differ significantly across municipalities.

The bootstrap only provides information about the bias in efficiencies that is due to sampling error, and we have also investigated whether the results from model A are robust to measurement errors and outliers, by performing jackknifing. Jackknifing means that we leave out each efficient municipality one at a time. Then we run a new DEA analysis in each case. In our case with 19 effective municipalities, 19 additional DEA analyses are run. When one efficient unit is left out, the mean efficiency score of the remaining units will generally increase.[11] The efficiency scores are considered to be robust if the increase is small. In our case the increase in mean efficiency is 1 percentage point or less in 17 of the 19 cases. In the last two cases the increase is slightly above 1 percentage point. Moreover, the rank correlation between

11 The mean efficiency for the remaining units is unaffected if the unit that is left out is not a reference for any ineffective unit.

the original efficiency scores and the various jackknife models varies between 0.97 and 1.

Table 4 provides more information about grades and teacher hours per student in the efficient municipalities (from model A), and they are also compared with other municipalities with roughly the same number of students. The efficient municipalities are divided into three groups; (i) municipalities with 45–85 students, (ii) municipalities with 190–400 students, and (iii) municipalities with 700–1400 students. The table does not include the four smallest efficient municipalities and the three largest efficient municipalities, as they cannot be compared with inefficient municipalities with a similar number of students.

The four efficient municipalities in the group with 45–85 students (EM1–EM4) are characterized by both high grades and low resource use per student. Their grades are on average 3%–7% above the mean of all other municipalities in the group, and the number of teacher hours per student is 22% below. In the group with 190–400 students the best-performing municipalities (EM5–EM8) are also characterized by high grades and low resource use. Grades are on average 1%–5% higher than the mean of the group, and teacher hours per student 22% lower. In the largest group (700–1400 stu-

Table 4
Describing the Efficient Municipalities (Model A)

	Adjusted assessment grades			Teacher hours per student
	Mathematics	Norwegian	English	
Municipalities with 45–85 students				
EM1 (48 students)	3.59	3.61	3.91	85.3
EM2 (53 students)	3.67	3.46	3.64	91.8
EM3 (80 students)	3.78	3.85	3.85	88.2
EM4 (82 students)	3.81	3.89	3.95	98.5
Mean for the rest of the group	3.57	3.58	3.59	117.7
Municipalities with 190–400 students				
EM5 (196 students)	3.52	3.53	3.50	64.3
EM6 (201 students)	3.91	3.56	3.64	69.0
EM7 (252 students)	3.91	3.56	3.64	69.0
EM8 (388 students)	3.57	3.74	3.53	61.4
Mean for the rest of the group	3.52	3.53	3.52	85.0
Municipalities with 700–1400 students				
EM9 (704 students)	3.40	3.34	3.37	63.2
EM10 (835 students)	3.56	3.54	3.39	62.5
EM11 (905 students)	3.40	3.46	3.51	61.3
EM12 (1375 students)	3.41	3.43	3.63	72.4
Mean for the rest of the group	3.53	3.52	3.53	84.3

dents) the efficient municipalities (EM9–EM12) do not have higher grades than the mean of the group, but come out as efficient because they have low resource use per student.

The characterization of the efficient municipalities is different from Duncombe, Miner, and Ruggiero (1997). They find that efficient New York State school districts are characterized by high resource use (expenditures per student) that pays off in high student performance. In the Norwegian case, efficient municipalities are rather characterized by low resource use, but it is not associated with low student performance.

Finally, we have briefly compared the results from our approach with the approach where the variables describing family background are included as inputs in the DEA analysis. When all available variables are included in the DEA analysis, around 80% of the municipalities come out as fully efficient. We take this as evidence that the approach cannot handle a large number of socioeconomic variables without producing too many efficient units. Even when we restrict the socioeconomic controls to gender, parents' education, immigrants, and students with special needs, around 50% of the municipalities come out as efficient.

6. Explaining Variation in Educational Efficiency

As noted by Worthington (2001, p. 265), efforts to explain variation in educational efficiency are underdeveloped. Most studies merely compare efficiency scores in different groups of the sample. Duncombe, Miner, and Ruggiero (1997) and Grosskopf et al. (2001) are two of the few studies that attempt to explain variation in educational efficiency. They focus on monitoring and competition between school districts. In this section we try to explain variations in educational efficiency along the lines of earlier studies of efficiency in Norwegian municipalities that focus on political and budgetary institutions. The earlier studies include Kalseth and Rattsø (1998), who analyze administrative spending, Kalseth (2003), who analyzes the care for the elderly sector, and Borge, Falch, and Tovmo (2005), who analyze all service sectors simultaneously.

With regard to political institutions, several studies of Norwegian municipalities have emphasized the importance of political strength. There is evidence that political strength contributes to lower user charges (Borge, 2000) and to lower budget deficits (Borge, 2005). One interpretation of these findings is that a strong political leadership has an advantage in opposing pressure from external interest groups to increase spending (which in turn has to be financed by higher user charges and/or higher budget deficits). Moreover, political strength reduces administrative spending (Kalseth and

Rattsø, 1998) and increases efficiency (Kalseth, 2003; Borge, Falch, and Tovmo, 2005), which indicates that a strong political leadership also has an advantage in opposing internal pressure to increase budgetary slack. A traditional Herfindahl index has been the most widely used indicator of political strength. The index is calculated as

$$HERF = \sum_{p=1}^{P} SH_p^2 , \tag{2}$$

where SH_p is the share of representatives from party p. The index takes the maximum value 1 when a single party holds all the seats in the local council; the minimum value $1/P$ is attained when the seats are equally divided among the P parties. The Herfindahl index is inversely related to the degree of party fragmentation in the local council and thereby positively related to strength. We expect the Herfindahl index to have a positive effect on efficiency.

In Norway the socialist camp is dominated by the Labor Party, while the nonsocialist camp is more fragmented. As a consequence, there is a positive correlation between the Herfindahl index and the share of socialists in the local council. Since we cannot rule out that socialist influence has an effect on efficiency, one could argue that it should be included in the analysis to get an unbiased estimate of the Herfindahl index. A more substantial argument is that earlier studies document that a large share of socialists is associated with high administrative spending (Kalseth and Rattsø, 1998), low efficiency in care for the elderly (Kalseth, 2003), and low overall efficiency (Borge, Falch, and Tovmo, 2005). A possible explanation for these results is that it might be harder for socialists to impose a hard budget constraint on service producers because they are more concerned about service quality

When it comes to budgetary procedures, we distinguish between centralized (top down) and decentralized (bottom up) procedures in the initial phases of the budget process. In a centralized budgetary procedure the head of the administration (for an administrative centralized procedure) or of the executive board (for a political centralized procedure) presents an overall budget proposal for each sector, and the sectors only work out specific details within themselves. In a decentralized or fragmented budgetary procedure, on the other hand, each sector works out its own budget proposals, while the head of the administration or the executive board coordinates an overall budget proposal to be approved by the local council. Tovmo (2006) finds that a centralized budgetary procedure contributes to lower budget deficits, while Borge, Falch, and Tovmo (2005) find no significant effect on overall efficiency. The budgetary variable is only available for 306 of the 426 municipalities included in this study.

The earlier studies (Kalseth and Rattsø, 1998; Kalseth, 2003; Borge, Falch, and Tovmo, 2005) also indicate that high levels of local government revenue are associated with low efficiency. The underlying argument may be that the service-producing agencies are able to take advantage of a rich sponsor to increase budgetary slack. As an indicator of municipal revenue we use local taxes and block grants per capita deflated by an index that captures varying cost conditions across municipalities. This revenue indicator is widely accepted as the most reliable indicator of differences in economic conditions across municipalities.

Existing evidence on the effect of school competition or school choice, either between public and private schools or between public schools, yields no clear-cut answers. Hoxby (2000) provides evidence that school choice induces greater school productivity in an analysis of U.S. metropolitan areas, whereas Fiske and Ladd (2001) find that parental choice has a negative effect on achievement in elementary schools in New Zealand. Most Norwegian municipalities apply a neighborhood school rule that limits the competition between public schools. However, some municipalities have introduced school choice. A recent survey, focusing on the introduction of new management tools, indicates that around 13% of the municipalities have introduced some form of choice between public schools. We construct a dummy variable for whether school choice is introduced or not. It is available for 245 of the 426 municipalities included in this study.

The determinants of educational efficiency are analyzed using tobit regressions. This is an appropriate method, since the dependent variable, the calculated efficiency score from the DEA analysis, is censored at 1. The regression results are presented in table 5, where we use the efficiency scores from the benchmark model A as dependent variable in the first six columns.[12] The first equation (I) disregards the possible effects of political and budgetary institutions discussed above, and is primarily a test of our method of controlling for family background. It includes the level of education in the municipality (measured as the fraction of the population with education above upper secondary level), the share of students with special needs, and the share of minority students. We follow earlier studies of educational efficiency in including average school size (linearly and squared). Finally, we control for structural differences between municipalities by including population size and the share of the population living in rural areas.

The level of education and the share of minority students come out as insignificant, which indicates that our approach of controlling for family

12 Table 10 reports descriptive statistics for the variables used in the tobit regressions.

Table 5
The Determinants of Educational Efficiency

Variable	I	II	III	IV	V	VI	VII	VIII
Level of education	−0.180							
	(1.26)							
Minority students	−0.386							
	(1.54)							
Students with	−0.884	−0.794	−0.772	−0.600	−0.780	−0.780	−0.790	−0.846
special needs	(5.15)	(4.76)	(3.64)	(2.90)	(4.37)	(4.31)	(4.42)	(4.77)
Average school	0.086	0.102	0.109	0.107				
size (in 100)	(4.39)	(4.65)	(4.01)	(3.94)				
School size squared	−0.008	−0.011	−0.013	−0.013				
	(1.75)	(2.24)	(2.14)	(2.13)				
Population size	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002
(in 1000)	(2.53)	(2.35)	(1.48)	(2.59)	(3.91)	(4.42)	(3.90)	(3.75)
Rural	0.025	0.018	0.000	0.024	−0.078	−0.078	−0.080	−0.091
	(0.89)	(0.68)	(0.01)	(0.73)	(3.05)	(2.93)	(3.13)	(3.61)
Municipal revenue		−0.042	−0.036	−0.033	−0.124	−0.124	−0.126	−0.129
		(1.50)	(1.07)	(1.07)	(4.54)	(2.22)	(4.60)	(4.77)
Herfindahl index		0.323	0.356	0.198	0.201	0.201	0.199	0.190
		(4.60)	(3.77)	(2.49)	(2.76)	(2.30)	(2.73)	(2.63)
Share of socialists		−0.142	−0.127	−0.091	−0.123	−0.123	−0.125	−0.125
		(3.61)	(2.46)	(1.97)	(2.93)	(2.74)	(2.97)	(2.99)
Centralized			−0.019					
budgetary proc.			(1.13)					
School choice				−0.019				
				(0.94)				
No. of observations	426	426	306	245	426	426	426	426
Log likelihood	308.8	319.5	220.5	204.1	293.1	293.1	290.7	292.5

Notes: Tobit estimates with absolute *t*-values in parentheses. The dependent variable in columns I–VI is the efficiency scores from model A. The efficiency scores from model B are used as the dependent variable in column VII, and the efficiency scores from model C in column VIII. The *t*-values reported in column VI are based on bootstrapped standard errors.

background and student characteristics works well.[13] On the other hand, the share of students with special needs comes out as significant, and a large share of students with special needs is associated with low efficiency. The significant effect of the share of students with special needs might reflect the lack of data on the individual level (see section 4). School size has a significantly positive effect on efficiency, while school size squared has a negative effect. However, the effect of school size does not reflect economies of scale, since variable returns to scale are allowed for in the underlying DEA analysis. It

13 Both variables come out as highly significant when they are regressed on the efficiency score from a DEA model where the raw assessment grades are used as outputs instead of the adjusted assessment grades.

rather reflects that the variation in efficiency across municipalities is related to average school size, and more precisely that the variation is larger among municipalities with small schools. Population size comes out with a positive and significant coefficient, while the share of the population living in rural areas is insignificant.

In column II we include municipal revenue and the political variables as additional explanatory variables. The two political variables come out highly significant. The positive sign of the Herfindahl index means that municipalities with highly fragmented local councils tend to have low educational efficiency. The share of socialists comes out with a negative sign, indicating that socialist influence is associated with low educational efficiency. Municipal revenue comes out with a negative coefficient, but it is not statistically significant.

The dummy variables for budgetary procedure and school choice are included as additional regressors in columns III and IV respectively.[14] Neither of the two variables comes out as significant. Moreover, the sign and significance of the two political variables and municipal revenue are unaffected when we control for budgetary procedure and school choice.

Our findings that party fragmentation and the share of socialists in the local council have a negative and significant effect on efficiency and that the choice of budgetary procedure is of little importance are in line with earlier studies of efficiency in Norwegian municipalities. However, in contrast to the earlier studies, the analysis does not produce a significant effect of municipal revenue. One might suspect that the reason for this discrepancy is that a high level of revenue is associated with a decentralized school structure, and that the effect of revenue is captured by school size in regressions I–IV. The suspicion is confirmed by regression V, where we exclude the school-size variables. Then municipal revenue comes out as highly significant and with the expected negative sign. In addition, the settlement pattern (the share of population in rural areas) becomes significant, and the quantitative effect of population size increases. Our understanding of these changes is that municipal revenue, population size, and settlement pattern represent background variables that explain school structure and average school size. As a consequence, the statistical significance and quantitative effects of these variables are reduced when actual school size is controlled for directly.

In the final three columns in table 5 we investigate the robustness of the results. The point of departure is column V, where the school-size variables are excluded. The first robustness test, reported in column VI, investigates

14 In order to limit the reduction in sample size, we do not include both variables in the same regression.

whether the significance of the explanatory variables is robust to bootstrapping of the standard errors. It appears that the *t*-values based on bootstrapped standard errors are very similar to those based on the usual standard errors. The main exception is municipal revenue, where the *t*-value is reduced by half, but the effect of municipal revenue is still significant at conventional levels of significance. In the final two columns we use the efficiency scores from models B and C respectively. It turns out that the sign and significance of the explanatory variables remain unchanged from column V, and also that the point estimates are reasonably stable across the specifications. In fact, this is no surprise, given the high correlation between the efficiency scores from the two models.

We finally illustrate the quantitative effect of municipal revenue and the two political variables using equation V in table 5. An increase in municipal revenue that amounts to 10% of the mean is expected to reduce educational efficiency by 1.2 percentage points. An increase in party fragmentation by 10 percentage points is predicted to reduce efficiency by 2 percentage points. The predicted effect of an increase in the share of socialists by 10 percentage points is to reduce efficiency by 1.2 percentage points. Compared to the analysis of overall efficiency by Borge, Falch, and Tovmo (2005), the quantitative effects of municipal revenue and party fragmentation are weaker in our case.

In table 6 we investigate how party fragmentation, the share of socialists, and municipal revenue affect efficiency. Does their negative effect on efficiency reflect high resource use, high student achievement, or both? The issue is investigated by running simple regressions with average adjusted grades and number of teacher hours per student as dependent variables, and with party fragmentation, the share of socialists, and municipal revenue as explanatory variables. In addition, the share of students with special needs is included in both equations. The inverse number of students and the share of the population living in rural areas are included in the teacher-hours-per-student equation.

The estimation results reveal that municipal revenue contributes to both high student achievement (assessment grades) and high resource use per student. However, we know from table 5 that the effect of increased resource use dominates the effect of higher achievement, i.e., a high level of municipal revenue contributes to low educational efficiency. For the political variables the effects on resource use and performance work in the same direction. A high degree of party fragmentation and a large share of socialists contribute to both increased resource use and lower achievement. The effect of party fragmentation is significant on achievement, but not on resource use. The share of socialists is highly significant in all equations except the equation for exam results.

Table 6
The Determinants of Adjusted Grades and Teacher Hours per Student

	Adjusted assessment grades in core subjects	Adjusted assessment grades in other subjects	Adjusted exam result	Number of teacher hours per student
Share of students with special needs	0.375 (1.58)	0.027 (0.12)	0.089 (0.29)	99.664 (3.95)
Municipal revenue	0.149 (4.08)	0.152 (4.33)	0.046 (0.98)	24.287 (5.90)
Herfindahl index of (inverse) party fragmentation	0.150 (1.62)	0.151 (1.71)	0.230 (1.94)	−11.783 (1.01)
Share of socialists in the local council	−0.102 (1.90)	−0.126 (2.44)	−0.105 (1.52)	16.258 (2.66)
Inverse number of students in local council				1.948 (16.36)
The share of the population living in rural areas				8.655 (2.59)
Observations	426	426	426	426
R^2	0.07	0.08	0.02	0.66

Notes: OLS estimates with absolute *t*-values in parentheses.

7. Concluding Remarks

The purpose of this paper was to calculate the efficiency potential in the lower-secondary-school sector in Norway and to analyze the efficiency variation across municipalities. In a DEA analysis, with grades adjusted for family background as outputs and teacher hours as inputs, the national efficiency potential was calculated to be 14%. The calculated efficiency potential is fairly robust to outliers and the formulation of the educational production function. Based on a comparison of municipalities with roughly the same number of students, we find that the efficient municipalities from the DEA analysis are characterized by relatively low resource use per student, and (except for the largest municipalities) they also have relatively high student achievement.

In a second-stage analysis we ran tobit regressions in order to explain the variations in efficiency scores across municipalities. We find that a fragmented local council, a high share of socialists, and a high level of municipal revenue are associated with low efficiency. In additional regressions we investigate how party fragmentation, the share of socialists, and the level of municipal revenue affect efficiency, i.e., whether the negative effect on efficiency reflects high resource use per student, low student performance, or both. For party fragmentation and the share of socialists we find that the negative effect on efficiency reflects both higher resource use per student and lower student performance. Higher municipal revenue contributes both to high student performance and to high resource use per student, but the overall effect is to reduce efficiency.

8. Appendix

Table 7

Descriptive Statistics for Grades in the Final Year

	Mean	Coefficient of variation	Min	Max
School year 2001/02				
Norwegian	3.81	0.060	3.00	4.83
English	3.69	0.072	2.73	4.73
Mathematics	3.44	0.081	2.29	4.45
Other subjects	4.02	0.049	3.33	4.78
Written exam	3.39	0.088	2.53	4.75
School year 2002/03				
Norwegian	3.83	0.065	3.00	4.75
English	3.69	0.065	2.83	4.56
Mathematics	3.46	0.078	2.67	4.33
Other subjects	4.04	0.049	3.74	4.76
Written exam	3.39	0.087	2.25	4.33

Notes: The figures are based on data for 426 municipalities. The reported means are unweighted averages.

Table 8

Descriptive Statistics for the Variables in the Student Level Regressions

Variable	School year 2001/02				School year 2002/03			
	Mean (s.d.)	Min	Max	No. of obs.	Mean (s.d.)	Min	Max	No. of obs.
Girl	0.472 (0.499)	0	1	51,098	0.4926 (0.500)	0	1	52,928
Immigrant	0.086 (0.280)	0	1	51,098	0.0626 (0.242)	0	1	52,928

Table 8
continued

Variable	School year 2001/02 Mean (s.d.)	Min	Max	No. of obs.	School year 2002/03 Mean (s.d.)	Min	Max	No. of obs.
Adopted	0.009 (0.095)	0	1	51,098	0.0090 (0.094)	0	1	52,928
Father's highest education is upper secondary	0.572 (0.495)	0	1	51,098	0.5751 (0.494)	0	1	52,928
Father's highest education is lower tertiary	0.179 (0.383)	0	1	51,098	0.1778 (0.382)	0	1	52,928
Father's highest education is upper tertiary	0.093 (0.290)	0	1	51,098	0.0884 (0.284)	0	1	52,928
Mother's highest education is upper secondary	0.581 (0.493)	0	1	51,098	0.5804 (0.493)	0	1	52,928
Mother's highest education is lower tertiary	0.257 (0.437)	0	1	51,098	0.2562 (0.437)	0	1	52,928
Mother highest education is upper tertiary	0.033 (0.178)	0	1	51,098	0.0337 (0.180)	0	1	52,928
Mean educational level of peers' parents*	7.884 (0.944)	2.4	12	49,756	7.890 (0.952)	3.4	12	50,663
Student born in second quarter	0.261 (0.439)	0	1	51,098	0.2603 (0.439)	0	1	52,928
Student born in third quarter	0.256 (0.436)	0	1	51,098	0.2539 (0.435)	0	1	52,928
Student born in fourth quarter	0.220 (0.414)	0	1	51,098	0.2257 (0.418)	0	1	52,928
Student born earlier than its cohort	0.009 (0.095)	0	1	51,098	0.0083 (0.090)	0	1	52,928
Student born later than its cohort	0.017 (0.127)	0	1	51,098	0.0165 (0.127)	0	1	52,928
Parents living together as married	0.644 (0.479)	0	1	51,098	0.6234 (0.485)	0	1	52,928
Parents are cohabitants	0.043 (0.204)	0	1	51,098	0.0488 (0.215)	0	1	52,928
Parents separated	0.032 (0.177)	0	1	51,098	0.0340 (0.181)	0	1	52,928
Parents divorced	0.093 (0.290)	0	1	51,098	0.0928 (0.290)	0	1	52,928
Single mother	0.193 (0.395)	0	1	51,098	0.1979 (0.398)	0	1	52,928
Father's income (in 100,000 NOK)	4.205 (5.570)	0	622.4	49,079	4.536 (9.883)	0	167.0	50,632
Mother's income (in 100, 000 NOK)	2.313 (1.805)	0	123.2	50,579	2.511 (3.407)	0	39.3	52,387
Share of students receiving adapted teaching at the school	0.068 (0.032)	0.435	1	1,102	0.069 (0.036)	0	1	1,079

* The educational level of peers' parents is measured on a scale from 0 to 8, where 0 is no schooling and 8 is education on PhD level. The variable reflects the sum of the level of education for both parents.

Table 9

The Determinants of Assessment Grades and Exam Results, Individual Level

	Norwegian	English	Mathematics	Other subjects	National exam
Girl	0.636 (89.42)	0.471 (59.08)	0.141 (18.22)	0.444 (74.64)	0.297 (32.62)
Immigrant	−0.157 (9.81)	−0.188 (10.21)	−0.262 (15.50)	−0.158 (11.46)	−0.229 (12.27)
Adopted	−0.258 (7.74)	−0.311 (8.28)	−0.575 (15.41)	−0.273 (10.78)	−0.420 (11.23)
Father's highest education is upper secondary	0.190 (21.84)	0.218 (22.27)	0.257 (25.26)	0.187 (25.06)	0.213 (20.29)
Father's highest education is lower tertiary	0.509 (45.98)	0.579 (45.63)	0.641 (47.09)	0.461 (48.00)	0.559 (41.05)
Father's highest education is upper tertiary	0.614 (43.06)	0.713 (42.65)	0.804 (45.64)	0.541 (45.92)	0.681 (39.82)
Mother's highest education is lower tertiary	0.254 (50.93)	0.276 (24.14)	0.318 (28.50)	0.535 (28.74)	0.613 (22.21)
Mother's highest education is lower tertiary	0.601 (27.42)	0.636 (46.00)	0.702 (51.23)	0.242 (52.03)	0.257 (44.55)
Mother highest education is upper tertiary	0.734 (38.82)	0.813 (37.45)	0.918 (40.64)	0.628 (42.32)	0.807 (36.05)
Student born in second quarter	−0.042 (5.30)	−0.029 (3.16)	−0.042 (4.35)	−0.027 (4.02)	−0.029 (3.05)
Student born in third quarter	−0.088 (11.50)	−0.068 (7.67)	−0.090 (10.02)	−0.074 (11.83)	−0.070 (8.03)
Student born in fourth quarter	−0.146 (17.82)	−0.123 (13.20)	−0.136 (13.56)	−0.118 (17.35)	−0.112 (12.02)
Student born earlier than 1986–1987	0.160 (5.83)	0.341 (10.34)	0.291 (8.41)	0.139 (6.15)	0.302 (8.69)
Student born later than 1986–1987	−0.419 (15.61)	−0.547 (18.06)	−0.496 (16.97)	−0.358 (16.78)	−0.532 (16.30)
Parents living together as married	0.257 (27.64)	0.230 (22.61)	0.383 (36.20)	0.316 (39.05)	0.246 (23.27)
Parents are cohabitants	0.144 (9.84)	0.138 (8.51)	0.229 (13.63)	0.179 (14.97)	0.146 (8.50)
Parents separated	0.056 (2.81)	0.020 (0.89)	0.093 (4.20)	0.059 (3.65)	0.052 (2.34)
Parents divorced	0.033 (2.08)	0.007 (0.39)	0.050 (2.77)	0.027 (2.05)	0.024 (1.32)
Single mother	−0.026 (1.71)	0.027 (1.65)	−0.035 (2.08)	−0.016 (1.25)	−0.010 (0.61)
Father's income	0.002 (1.77)	0.002 (1.90)	0.003 (1.75)	0.002 (1.65)	0.002 (1.58)
Mother's income	0.006 (2.39)	0.008 (2.84)	0.009 (2.58)	0.008 (3.01)	0.006 (2.16)
Share of students receiving adapted teaching at the school	0.044 (0.22)	−0.385 (1.79)	−0.513 (2.31)	−0.206 (1.20)	−0.371 (1.78)

Table 9
continued

	Norwegian	English	Mathematics	Other subjects	National exam
Written exam in mathematics					−0.234 (19.41)
Written exam in Norwegian					0.083 (6.16)
Mean educational level of peers' parents					0.040 (5.06)
No. of observations	98,647	97,484	98,501	99,311	94,424
R^2	0.27	0.21	0.21	0.29	0.20

Notes: OLS estimates with absolute *t*-values in parentheses. Municipal fixed effects and year-specific intercepts (not reported) are included in all equations.

Table 10
Descriptive Statistics for the Variables in the Tobit Regressions

Variable	Mean	Standard deviation	Minimum value	Maximum value
Level of education	0.167	0.049	0.079	0.422
Share of minority students	0.027	0.023	0	0.184
Share of students with special needs	0.073	0.031	0	0.176
Average school size (in 100)	1.546	1.010	0.110	4.826
Share of the population living in rural areas	0.486	0.267	0.004	0.997
Population size (in 1000)	10.55	30.18	0.35	512.09
Municipal revenue	1.039	0.214	0.880	3.180
Herfindahl index of (inverse) party fragmentation	0.266	0.087	0.140	1
Share of socialists in the local council	0.366	0.141	0	0.846
Centralized budgetary procedure	0.831	0.373	0	1
School choice	0.27	0.333	0	1

References

Banker, R. D., Charnes, A., and Cooper, W. W. (1984), Some Models for Estimating Technical Scale and Inefficiency in Data Envelopment Analysis, Management Science 30, 1078–1092.

Bessent, A., Bessent, W., Kennington, J., and Reagan, B. (1982), An Application of Mathematical Programming to Assess Productivity in the Houston Independent School Districts, Management Science 28, 1355–1367.

Bonesrønning, H., and Rattsø, J. (1994), Efficiency Variation among the Norwegian High Schools: Consequences of Equalization Policy, Economics of Education Review 13, 289–304.

Borge, L.-E. (2000), Charging for Public Services: The Case of Utilities in Norwegian Local Governments, Regional Science and Urban Economics 30, 703–718.

Borge, L.-E. (2005), Strong Politicians, Small Deficits: Evidence from Norwegian Local Governments, European Journal of Political Economy 21, 325–344.

Borge, L.-E., Falch, T., and Tovmo, P. (2005), Efficiency in Public Service Production: The Impact of Political and Budgetary Institutions, Mimeo, Department of Economics, Norwegian University of Science and Technology.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978), Measuring the Efficiency of Decision Making Units, European Journal of Operational Research 2, 429–444.

Coelli, T., Prasada Rao, D. S., and Battese, G. (1998), An Introduction to Efficiency and Productivity Analysis, Kluwer Academic Publishers.

De Borger, B., and Kerstens, K. (1996), Cost Efficiency of Belgian Local Governments: A Comparative Analysis of FDH, DEA and Econometric Approaches, Regional Science and Urban Economics 26, 145–170.

Duncombe, W., Miner, J., and Ruggiero, J. (1997), Empirical Evaluation of Bureaucratic Models of Inefficiency, Public Choice 93, 1–18.

Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (2001), On the Determinants of School District Efficiency: Competition and Monitoring, Journal of Urban Economics 49, 453–478.

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

Hanushek, E. (2002), Publicly Provided Education, Working Paper 8799, National Bureau of Economic Research.

Hjalmarsson, L., Kumbhakar, S. C., and Heshmati, A. (1996), DEA, DFA and SFA: A Comparison, Journal of Productivity Analysis 7, 303–327.

Hoxby, C. M. (2000), Does Competition Among Public Schools Benefit Students and Taxpayers?, American Economic Review 90, 1209–1238.

Hægeland, T., Raaum, O., and Salvanes, K. G. (2004), Pupil Achievement, School Resources and Family Background, Mimeo, Ragnar Frisch Centre for Economic Research and Statistics Norway.

Kalseth, J. (2003), Political Determinants of Efficiency Variation in Municipal Service Production: An Analysis of Long-Term Care in Norway, PhD thesis, Department of Economics, Norwegian University of Science and Technology.

Kalseth, J., and Rattsø, J. (1998), Political Control of Administrative Spending: The Case of Local Governments in Norway, Economics and Politics 10, 63–83.

Kirjavainen, T., and Loikkanen, H. A. (1998), Efficiency of Finnish Senior Secondary Schools: An Application of DEA and Tobit Analysis, Economics of Education Review 17, 377–394.

Ladd, H. F., and Fiske, E. B. (2003), Does Competition Generate Better Schools? Evidence from New Zealand, Working Paper SAN01–16, Terry Sanford Institute of Public Policy, Duke University.

Simar, L., and Wilson, P. W. (1998), Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models, Management Science 44, 49–61.

Simar, L., and Wilson, P. W. (2000), Statistical Inference in Nonparametric Frontier Models: The State of the Art, Journal of Productivity Analysis, 12, 49–78.

Simar, L., and Wilson, P. W. (2004), Performance of the Bootstrap for DEA Estimators and the Iterating Principle, in: Cooper, W. W., Seiford, L. M., and Zhu, J. (eds.), Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston, 265–298.

Tovmo, P. (2006), Budgetary Procedures and Deficits in Norwegian Local Governments, Economics of Governance, forthcoming.

Worthington, A. C. (2001), An Empirical Survey of Frontier Efficiency Measurement Techniques in Education, Education Economics 9, 245–268.

Lars-Erik Borge
Department of Economics & Center for Economic
Research
Norwegian University of Science and Technology
Dragvoll Campus
7491 Trondheim
Norway
Email: larseb@svt.ntnu.no

Chapter 3
Teacher Hiring Practices and Educational Efficiency

Teacher Hiring Practices and Educational Efficiency

Linn Renée Naper*

Abstract

This paper investigates how teacher hiring practices affect educational efficiency in Norwegian school districts. The hiring decision is made at the school level by the principal or at the school district level. According to the data, efficiency is the highest in districts where hiring is decentralized. Hiring practices are decided by the school district, and linear estimates of the effect of decentralized hiring on efficiency may be biased because of non-random selection. First, I approach this problem by including a large set of controls in a school district level analysis, which does not alter the qualitative result. Second, I perform a school level analysis with district fixed effects. The results indicate, as expected, that the effect of decentralization is stronger for schools facing excess teacher supply than for schools without excess supply.

Keywords: Education, efficiency, DEA, school autonomy, hiring practices.
JEL classifications: I21, I29

1. Introduction

School sector performance has received significant attention in recent years and there is a constant search for reforms that may improve achievement and efficiency. There has been increasing interest in decentralized decision making within the education sector, and actions have been taken to increase school autonomy in the belief that relieving schools from central regulation will result in better school management and increased educational efficiency.

This paper analyses how teacher hiring practices affect educational efficiency in Norwegian schools districts. A decentralized hiring practice means that the appointment decision is made at the school level by the principal, whereas a centralized hiring practice means that the decision is made by the school district administration.

* Helpful comments and suggestions from Lars-Erik Borge, Torberg Falch, Bjarne Strøm, Jørn Rattsø, Hans Bonesrønning, Hessel Oosterbeek, Tarek Mostafa and seminar participants at The World meeting of the Public Choice Society in Amsterdam 2007 and the Norwegian University of Science and Technology are gratefully acknowledged.

The design of the hiring process may influence the composition of teachers because the judgements of who are the best applicants may differ across decision makers. Teachers are considered to be important to educational outcomes, but formal qualifications, such as education and experience, are generally not regarded as sufficient indicators of teacher effectiveness (Rockoff, 2004; Rivkin et al., 2005). Informal characteristics of teachers, including personal chemistry with students and colleagues, student composition, and the schools' working environment, may also be important. Informal qualifications are hard to observe in written applications, but are much more easily revealed in an interview situation. If appointment decisions are partly based on the informal characteristics of the applicants, factors such as local information and knowledge of the school are likely to play an important role to find the best teacher for a given position. School principals may know better than the school district administration which qualifications are most needed for a given teaching position and may therefore more effectively evaluate the applicants. Decentralized hiring practices may as a consequence result in an overall better match between schools, teachers, and students and are in this case expected to improve educational efficiency. However, with no extra precautions, increased decision-making authority for school principals may involve reduced school district control within the appointment process. Reduced control may allow for a more subjective evaluation of teacher candidates. If school principals make allowance for personal or social factors, for example, by employing a familiar teacher or previous colleagues, the objectively best candidates may not be hired and the appointment process may result in an overall less favourable match between teachers and students. In this case decentralization is expected to harm educational efficiency.

Several institutional features of a school system can be grouped under the heading of autonomy or decentralization. The existing literature suggests that increased autonomy within the educational sector may have positive effects on educational outcomes. In a cross-country analysis, Wößmann (2003) shows that students tend to perform better in countries where schools have autonomy in personnel and day-to-day decisions, in particular when schools are held accountable for results. However, it is challenging to produce evidence on the impact of autonomy because the degree of decentralization, generally, is a national decision and therefore does not vary much within countries. Positive effects of autonomy in school management are also reported by country-specific studies such as Clark (2005) for the UK and Eskeland and Filmer (2007) for Argentina. The theoretical contribution by Bishop and Wößmann (2004) indicates that decentralization is more efficient when applied to areas where school-specific information is important and where the prospect for school level opportunistic behaviour is limited. The combined effect of increased autonomy and financial incentives is analysed empirically by Hoxby and Rockoff (2004). They report positive effects from US charter schools where increased autonomy in school management is combined with competition for students.[1]

Empirical evidence is relatively scarce when it comes to analyses of teacher hiring practices, and existing papers have taken different approaches. Some studies focus on how teacher composition is related to local hiring practices, while others are concerned with how hiring practices affect educational performance (efficiency or student achievement). Ballou (1996) presents evidence that public schools do not hire the best applicants for vacant teaching positions in an analysis of the recruitment process of prospective teachers in the US. Balter and Duncombe (2006) study the hiring process in New York school districts, and find that some features of the process, such as recruitment incentives and extra-curricular activities, are associated with more qualified teachers. These studies do not provide information on how hiring practices affect school performance, and the empirical conclusions and implications of these studies strictly depend upon the assumption made regarding the definition of teacher quality.

In a cross-country analysis based on PISA 2000, Robin and Sprietsma (2003) find that school autonomy in decisions regarding the recruitment of teachers is positively related to student achievement levels. Furthermore, in a review of mainly educational research on ‘effective schools’, Zigarelli (1996) presents evidence that principal influence in hiring and firing of teachers is one of three variables characterizing ‘effective schools’. This paper also focuses on the effect of hiring practices on educational performance and adds to the literature with an empirical analysis of variation in hiring practices in Norway. The available data clearly provide an advantage over earlier contributions on this topic.

Information about hiring practices is collected in a large survey on organizational approaches in Norwegian local public service provision. The information is at the school district level, and indicates whether the school district has decentralized the hiring decision to individual schools or not. The dependent variable in the empirical analysis is a measure of educational efficiency, measured by data envelopment analysis (DEA). The DEA model is based on the same set-up as proposed by Borge and Naper (2006). The effect of hiring practices on

[1] The evidence on charter schools is mixed. See, for example, Hanushek et al. (2007) and Bettinger (2005).

efficiency is initially investigated in a school district Tobit model accounting for school district heterogeneity by including a rich set of observable school district characteristics. There is a predominance of large school districts that decentralize and larger school districts are also found to be more efficient. The empirical analysis therefore particularly explores the robustness of the estimated effects related to the size of the school district. The empirical results suggest that school districts where the principals of individual schools are given the responsibility of hiring teachers are more efficient than school districts with traditional centralized hiring practices. This effect does not seem to be driven by the variation in school district size.

There is no available information about factors that trigger the decision to decentralize and the causal effect of hiring practices may be hard to pin down. If the decision to decentralize is based on unobservable school district characteristics that are correlated with educational efficiency, the Tobit estimates may be biased. In a second approach, I therefore allow for heterogeneous treatment effects for individual schools in a school level analysis exploiting that the impact of decentralized hiring is expected to be stronger when schools have excess teacher supply. By including school district fixed effects I can estimate the interaction effect between decentralized hiring and teacher supply, although not the direct effect of decentralized hiring, using within school district variation only. Given that hiring practices matter for the overall efficiency of a school district as reported from the Tobit, we should indeed expect to find the strongest effects for schools with an actual option between candidates. The empirical analysis produces a positive and significant interaction effect, and therefore provides support for this hypothesis.

The remainder of this paper is organized in the following way. Section 2 discusses some relevant features of the institutional setting in Norway. Section 3 presents the data. Section 4 discusses the empirical specification, Section 5 presents the empirical results, and Section 6 offers a brief summary and some concluding remarks.

2. Institutional Setting

Compulsory school in Norway is 10 years and students enrol at primary school (1^{st} to 7^{th} grade) in the year they turn six and finish lower secondary school (8^{th} to 10^{th} grade) the year they turn 16. Most primary and lower secondary schools are owned and operated by the municipalities. Private schools account for less than 2 per cent of the students. Primary schools are typically smaller than lower secondary schools and most municipalities therefore operate more primary schools than lower secondary schools. The lower secondary schools either have students only at the lower secondary level or are mixed schools with students at both primary and lower secondary levels. Mixed schools typically also receive students from several primary schools at the lower secondary level. The municipalities are multi-purpose authorities which, in addition to compulsory education, are responsible for welfare services such as child care, primary health care, and care for the elderly. Other important tasks are culture and infrastructure.

The financing of municipalities is quite centralized and the main revenue sources are taxes (43%), block grants (21%), earmarked grants (13%), and user charges (16%). Around 95 per cent of local taxes are regulated income and wealth taxes (where all municipalities apply the maximum rates). The opportunity to influence current revenues is limited to property taxes and user charges. However, the municipalities enjoy more discretion on the spending side than on the revenue side, and the allocation of taxes and block grants between different service sectors is decided locally. There are few earmarked grants and user charges in the education sector, so education is mainly financed by taxes and block grants. Each municipality represents a school district and in the remainder of this paper I will use the term 'school district' when referring to municipalities.

The education sector is regulated by the Education Act of 1992[2] and Norway is traditionally considered to have a strict centralized system. Official OECD publications list Norway among the countries with the least autonomous schools together with countries such as Greece and Portugal (OECD, 1998). Nevertheless, since the late 1980s there has been a trend towards more decentralization and delegation of decision-making power from the central government to school districts, and from schools districts to schools (Lyng and Blichfeldt, 2003). National guidelines allow school districts to adopt different ways of exercising their authority and organizing the management of schools. School districts have generally chosen either a centralized approach where most decisions on school management are taken at the district level or a decentralized approach where decision-making power is delegated to individual schools. Decentralization means that various financial and personal-administrative tasks are delegated to the school principal(s), often by closing down the 'municipal school office' (Skolekontor). The principal then answers directly to the district's chief executive officer

[2] See www.lovdata.no for information on the laws and guidelines in the Education Act.

(Rådmann). Centralized school management also involves the principals, but the final decisions are taken at the district level and the 'municipal school office' is responsible and answers to the chief executive officer.

The appointment process of teachers is one of the areas where decentralization has become widespread. Although hiring of teachers is formally the responsibility of the school districts, several districts have chosen to delegate the decision to the school principal(s). The decision to decentralize always concerns all schools within the school district, so we can only observe variation in decentralization across school districts. Independently of the chosen hiring practice, vacant teaching positions are always announced in collaboration with the school district. However, in school districts that decentralize, the applicants are asked to send the application directly to the school, emphasizing that it is the principal who is responsible in the process, including interview rounds, evaluation, and appointment of applicants. The process is always closely monitored by the teacher union and representatives from the union must be informed prior to every appointment decision. An applicant without a teacher certification can only be hired if no applicants satisfy the formal qualification requirements. Hiring of non-certified teachers is always temporary and the position should be announced again within one year. There are, however, no formal instructions on how to select or prioritize between certified applicants. The appointment decision may therefore vary depending on whether the decision is made by the school district or by the school.

When it comes to determining teacher wages and workload the school districts have less influence on the process. Wages were determined in national bargains between teacher unions and the central government until the school year 2000–2001. In 2001 some limited local flexibility was introduced, but wages are still mainly determined by teachers' formal education and experience. The national contract therefore effectively prevents school districts from attracting teachers by means of wage policy or regulation of the workload.

3. Data and Descriptive Statistics

3.1 Hiring practices and educational efficiency

Information on hiring practices is collected in a large survey on organizational approaches in local public service provision[3]. The questionnaires include questions based on the 'New

[3] The survey is rendered by the Institute for Research in Economics and Business Administration (Bergen) and Centre for Economic Research (Trondheim).

Public Management' tradition.[4] The survey is part of a project on efficiency in local public service provision in Norway, and in addition to the educational sector, information was collected for child care and care for the elderly. The questionnaire was sent to the chief executive officer in all 433 school districts in autumn 2004 and was completed during spring 2005. Overall the response rate was close to 65 per cent. However, some districts only partly responded to the questionnaire and for the question regarding hiring practices the data provide information for 58 per cent (250) of the school districts.[5] The empirical analysis also includes economic, demographic, and political information about schools and school districts provided by The Norwegian Social Science Data Service and Statistics Norway.

The dependent variable used in the empirical analysis is a measure of educational efficiency calculated using DEA, which uses the observed input–output bundles to estimate a best practice reference production frontier by linear programming. Efficiency is calculated for both the school and the school district level in two different calculations. The efficiency score for each unit (school or school district) is calculated as the relative distance to the efficiency frontier.[6] The scores vary between zero and one, and a score equal to one indicates that the unit is fully efficient and operates at the efficiency frontier. The DEA model applied in this paper includes one input variable and three output variables and is based on the same approach as proposed by Borge and Naper (2006). The model allows for variable returns to scale, and I have chosen to focus on input-oriented efficiency indicators.

The input variable included in the DEA model is the average number of teacher hours per student for the two academic years 2002–2003 and 2003–2004. The output variables are based on individual teacher-assessed marks in three subjects, Mathematics, English, and Norwegian, for the same years. Prior to the DEA analysis, the marks are adjusted for observable differences in student and family characteristics by regressing individual student achievement in each subject on a set of individual student characteristics. Information about individual students is provided by Statistics Norway and includes students' gender, age, and immigration status, together with parental marital status, income, and education. The regressions include year and unit fixed effects. It is the estimated unit fixed effects, and not

[4] See, for example, Christensen and Lægreid (2005) for more information on New Public Management theories.
[5] The survey also includes questions on decentralization of responsibilities within other managerial areas, but in this paper I chose to focus on teacher hiring practices. The Norwegian report Haraldsvik and Naper (2007), however, shows that the results presented in the present paper are robust to the inclusion of variables measuring decentralization along other dimensions.
[6] See Charnes et al. (1978) for more on DEA and also Worthington (2001) for a review on applications of DEA to the education sector.

the raw marks, that are included as output variables in the DEA model. The school and school district fixed effects are interpreted as the average student achievement level, in the school or school district, respectively, adjusted for observable student characteristics and family background.

Table 1
DEA efficiency scores for individual schools and school districts

	School districts	Schools
No. of units	250	653
Mean	0.71	0.66
Minimum	0.42	0.33
1st quartile	0.62	0.54
3rd quartile	0.80	0.76
Share of efficient units (%)	3.2	1.2

Note: Mean efficiency values and indicators of variation for individual schools and school districts.

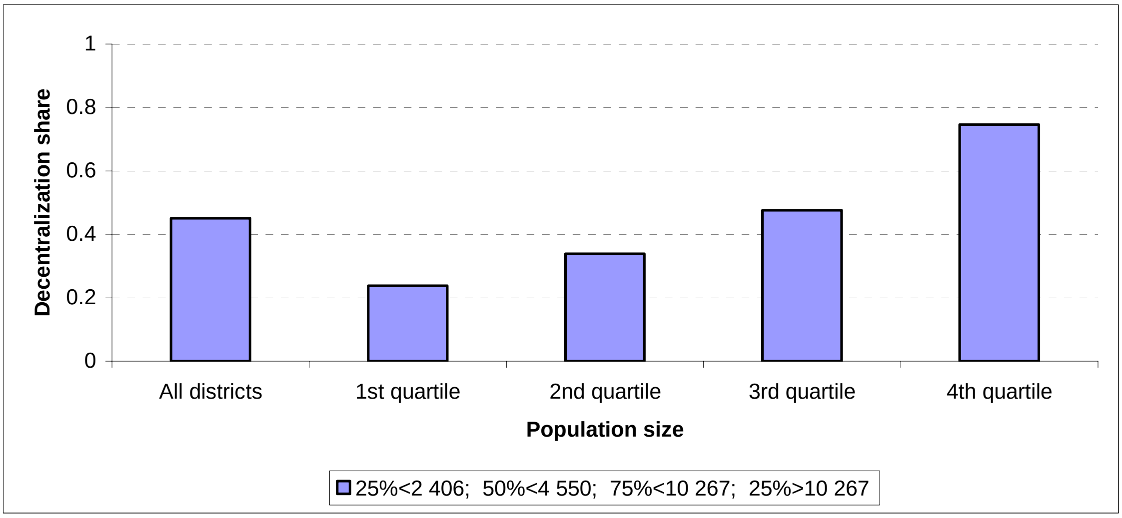
Table 1 presents the results from the two DEA analyses. For school districts the mean efficiency is 71 per cent, meaning that the average school district in the sample could reduce input by 29 per cent without reducing outputs if it operated at the efficiency frontier. By looking at the minimum efficiency level and the 1st and 3rd quartile sample values, we observe a substantial variation in efficiency across both school districts and individual schools.

As expected we find that the mean efficiency score and the share of efficient units are lower when the DEA model is calculated for individual schools. One characteristic of the DEA method is that the number of efficient units and the calculated efficiency potential depend on the number of inputs and outputs relative to the sample size. For a given number of variables an increase in the sample size will therefore reduce mean efficiency.

3.2 School district size

Figures 1a and 1b below illustrate the distribution of hiring practices and educational efficiency by school district size. School district size is measured by population size. In Figure 1a the share of districts with a decentralized hiring practice is presented for all districts together and separately for each population quartile. As illustrated by the left bar in the diagram, around 45 per cent, or 113 school districts, report to have chosen a decentralized hiring practice. Furthermore, the figure reveals that the variation in hiring practices is related to school district size. Close to 75 per cent of the school districts in the 4th quartile of the distribution have a decentralized hiring practice, while the share is 24 per cent in the 1st quartile. Among the largest districts only decentralized hiring is observed. Decentralization is present in all parts of the distribution, but there is a predominance of large school districts that chose to decentralize.

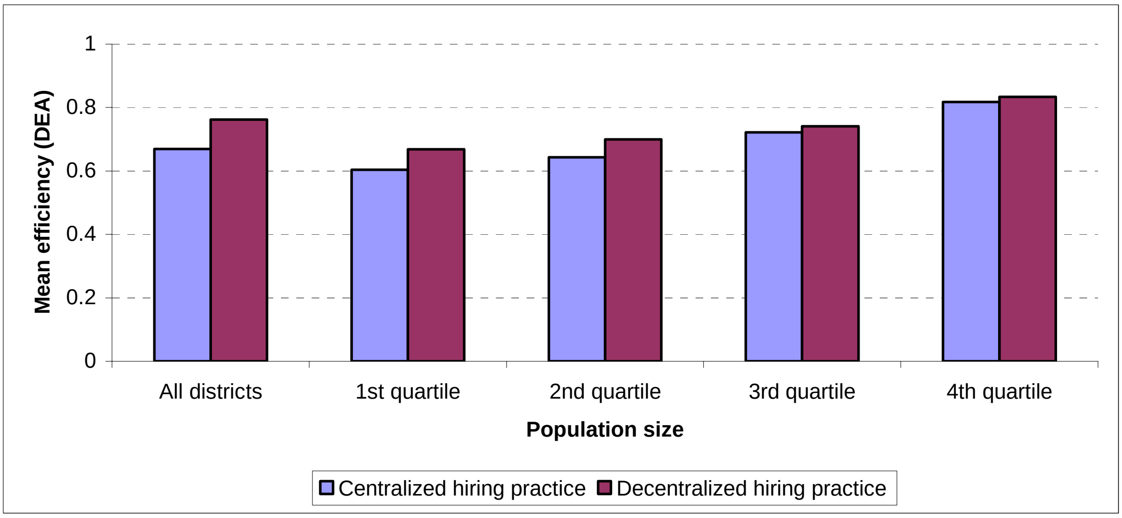
Figure 1a
Hiring practices and school district size



In Figure 1b, mean efficiency levels (for school districts) are presented separately for school districts with a centralized and a decentralized hiring practice. The two left bars in the diagram compare mean efficiency for all school districts and show that efficiency on average is about 9.5 percentage points higher for those school districts that decentralize than for others. A comparison of efficiency scores for individual schools in the two groups of districts shows that the difference in mean efficiency is also close to 10 percentage points in favour of individual schools in districts that decentralize.

Next, Figure 1b compares mean efficiency levels for school districts within each population quartile and some interesting features of the data are revealed. First, efficiency increases gradually with district size. Second, the differences in efficiency across teacher hiring practices are larger for smaller school districts than for larger districts. The difference in mean efficiency is roughly 6 percentage points for school districts in the lower half of the distribution and less than 2 percentage points for school districts in the upper half of the distribution.

Figure 1b:
Hiring practices and school district level efficiency



That school district size is correlated both with the decision to decentralize and with educational efficiency introduces an empirical challenge because the estimated effect of hiring practices may be confounded with the effect of school district size. In the empirical analysis in Section 5 it is therefore very important to include proper controls for school district size.[7] The average difference in efficiency over the four quartiles is close to 5 percentage points. Therefore, controlling for observable variation in school districts size, the mean difference in efficiency is reduced by roughly 50 per cent. As discussed in the introduction of the paper, I will also address the problem with potential non-random selection by allowing for heterogeneous treatment effects related to teacher supply.

3.3 Teacher supply

An obvious measure of teacher supply is the share of applicants that do not receive a job offer at each school, but this information is not available. However, we do have information on the share of certified teachers employed at each school. This variable may serve as a proxy for teacher supply (Bonesrønning et al., 2005). The intuition for this approach is that the use of uncertified teachers implies excess demand because schools are obliged to place priority on certified teachers. It is also important for this approach that teachers' wages are mainly determined nationally and therefore schools cannot attract teachers by raising wages. Strictly speaking we do not know whether schools that employ only certified teachers have excess

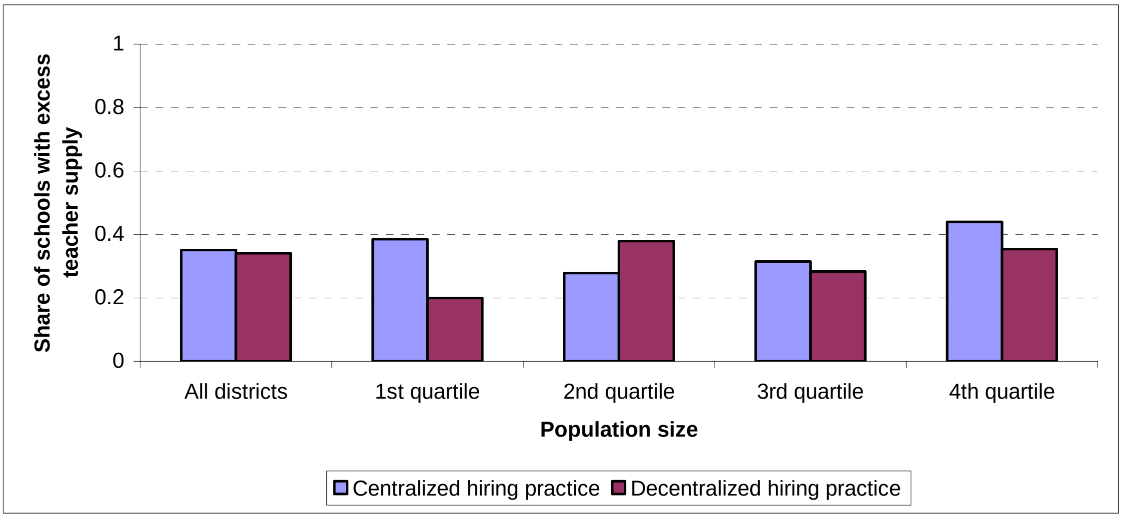
[7] In Table A1 of the appendix, the full list of variables used in the empirical analyses is presented, with both school district and individual school characteristics.

supply or if supply equals demand for these schools. For simplicity I will refer to schools with only certified teachers as schools with excess teacher supply in the remainder of this paper.

In the empirical analysis the focus is on separating excess demand and excess supply. Teacher supply is therefore included by using a dummy variable indicating whether schools employ only certified teachers in two subsequent years. It is reasonable to use persistence over time because yearly turnover among certified teachers in Norway is about 10 per cent. However, decentralization is a relatively recent phenomenon and the timing of decentralization most likely varies across school districts. Using lagged data may involve using variation in supply that occurred prior to the organizational change. Therefore, the measure is limited to the same year and the year prior to the survey.

The data reveal substantial variation in this measure of excess teacher supply, both within school districts and across school districts. Only 14 districts have excess teacher supply in all schools over the two years. For these districts the number of schools varies from two to five.[8] In Figure 2 the share of schools with excess teacher supply is presented separately for school districts with centralized and decentralized hiring practices. The two left bars in the diagram show that the share of schools with excess teacher supply is about 35 per cent for districts with centralized hiring practices and slightly lower for districts with decentralized hiring practices.

Figure 2
School level teacher supply by hiring practice



[8] School districts with only one lower secondary school are excluded from the sample in the district fixed effect models estimated in Section 5.

When school districts are sorted by population size, this picture changes. Within the 1st quartile the share of schools with excess teacher supply is close to 40 per cent for school districts with centralized hiring practices and only 20 per cent for districts with decentralized hiring practices. The share of schools with excess teacher supply is close to 10 percentage points higher for districts with centralized hiring practices when considering the districts in the 4th quartile. The exception is the 2nd quartile, where the share of schools with excess teacher supply is higher for school districts with decentralized hiring practices. The difference in teacher supply between schools in districts with centralized and decentralized hiring practices is therefore larger when school district size is controlled for.

4. Empirical Specification

Inference on the impact of institutional settings within a country requires variation in implementation, either across time or across units. In this paper I exploit the fact that some school districts in Norway have chosen to delegate hiring responsibilities to individual schools, while others have maintained a traditional centralized hiring practice. The cross-sectional variation in the variable of interest is a result of local adaptations to the existing legal framework. This introduces an econometric challenge because the decision to decentralize may not be random. Furthermore, if the factor(s) that triggers the decision is unobserved it is hard to pin down the causal effect of hiring practices and naïve estimates may suffer from an endogeneity bias.

I approach the potential endogeneity problem in two ways. First, in equation (1) observable school district heterogeneity is controlled for. The efficiency score e_i of school district i is regressed on the dummy D_i for decentralized hiring together with a rich set of observable school district characteristics represented by the vector M_i. The parameter μ_i is an assumed white noise error term.

(1) $$e_i = \alpha + \delta_1 D_i + \delta_2 M_i + \mu_i$$

The DEA efficiency scores are censored at one and equation (1) is therefore estimated by Tobit.[9] Assuming that the selection of hiring practices are based on observable school district characteristics, the Tobit estimate of the effect of hiring practices on school district efficiency is unbiased.

[9] This is common practice in the DEA literature (see, for instance, Worthington, 2001).

It is documented in Section 3 that school district size is correlated with both the decision to decentralize and educational efficiency. Equation (1) should therefore include a control for school district size. In the empirical analysis in Section 5 school district size is measured by the number of students enrolled in compulsory schools[10] and included in log and squared log form. The analysis of the determinants of efficiency in Borge and Naper (2006) indicates that a high level of revenues, a high degree of party fragmentation, and a high share of socialists in the local council is associated with low efficiency. These variables are therefore potential important control variables and are included in equation (1). [11] As in Borge and Naper (2006) I also include a variable describing the settlement pattern in the school district and the share of students with special needs. In addition I include the share of highly educated people[12] in the school district. This is potentially an important control variable because the educational level of the electorate may affect the organizational approach taken by the school district, and a high level of education in the school district may reflect higher student performance and more efficient schools.

If the decision to decentralize is based on unobservable school district characteristics that are correlated with educational efficiency, the Tobit estimate will be biased. For example, Tobit will underestimate the effect of hiring practices if a decentralized practice typically is chosen by relatively inefficient school districts to improve their performance. However, if school districts that decentralize tend to have well-managed and efficient schools, the effect will be overestimated. One way to overcome this problem, and to obtain unbiased estimates, is to use an instrumental variable (IV) that is correlated with the decision to decentralize but has no direct effect on efficiency. The idea behind IV estimation is to identify the effect using strictly exogenous variation. Alternatively, the effect could be identified using within school district variation in a district fixed effects (FE) model. The idea behind FE estimation is to identify the effect using only within school district variation. The included FEs absorb the effect of all unobserved district specific variables that influence the decision to decentralize.

The question is whether one or both of these methods can be used to identify the effect of hiring practices in equation (1). First, educational efficiency is the result of various local

---

[10] The number of students is highly correlated with population size and the empirical results in Section 5 are similar if population size is used instead of the number of students.
[11] The socialist share is defined as the share of representatives in the local council from the socialist camp, which is dominated by the Labor Party (AP) but also includes representatives from three other parties SV, NKP, and RV.
[12] The educational level is measured by the fraction of the population with education beyond the upper secondary level.

decisions and priorities within the education sector, as well as within other responsibility areas of the local authority (child care, health care, infrastructure, etc). It is therefore very difficult to find a variable that influences the decision to decentralize and at the same time is uncorrelated with efficiency. Besides, regressing the dummy for decentralized hiring on available school district characteristics, few variables, except school district size, turn out to be correlated with hiring practices. Using an IV is therefore not a feasible alternative. Second, FE estimation requires variation either over time or within school district, but the available information on hiring practices is only measured in one year and at the school district level. Therefore, it is not possible to identify the effect of hiring practices from equation (1) using FE estimation.

I therefore propose a slightly different approach where within school district variation is exploited by allowing for heterogeneous treatment effects for individual schools. I exploit the fact that decentralized hiring is expected to differ most from centralized hiring when schools have excess teacher supply, and define a school level model where the effect of decentralized hiring D_i interacts with the dummy S_{ji} for excess teacher supply for school *j* in school district *i*. The model is given by equation (2):

(2) $$e_{ji} = \alpha + \gamma_1 D_i + \gamma_2 D_i S_{ji} + \gamma_3 S_{ji} + \gamma_4 E_{ji} + \gamma_5 M_i + \mu_{ji},$$

where e_{ji} is the efficiency of school *j* in school district *i*, E_{ji} is a vector of individual school characteristics, M_i a vector of school district characteristics, and $\mu_{ji} = \eta_i + \varepsilon_{ji}$ is a composite error term. From this model the effect of decentralized hiring on efficiency is given by:

(3) $$\frac{\partial e_{ji}}{\partial D_i} = \gamma_1 + \gamma_2 S_{ji}.$$

The effect consists of two terms: (i) the direct effect of decentralized hiring on efficiency γ_1 and (ii) the interaction between excess supply and decentralized hiring γ_2. Individual schools in school districts where hiring is decentralized are generally expected to make better matches between teachers, schools, and students, ($\gamma_1 \succ 0$). Furthermore, this effect is expected to be stronger for schools facing excess teacher supply, ($\gamma_2 \succ 0$). Therefore, I expect the interaction effect γ_2 to be positive. Equation (2) may be estimated using school district FE. The model is then reduced to:

(4) $$e_{ji} = \alpha_i^* + \gamma_2 D_i S_{ji} + \gamma_3 S_{ji} + \gamma_4 E_{ji} + \varepsilon_{ji}, \quad \alpha_i^* = \alpha + \gamma_1 D_i + \gamma_5 M_i + \eta_i \,.$$

In equation (4) the school district specific constant term α_i^* captures the effect on efficiency of all school district characteristics, including the direct effect γ_1 of decentralized hiring and unobserved school district specific factors that affect the decision to decentralize. By exploiting within school district variation only, the direct effect γ_1 cannot be identified. However, the district FE estimation provides an estimate of the interaction effect γ_2. Although the total effect of hiring practices cannot be measured precisely, the estimated interaction effect provides useful information about the robustness of the effect found in the Tobit estimation. If teacher hiring practices matters for overall efficiency in a school district, we should expect to find the strongest effects for schools with an actual option between teacher candidates.

It should be kept in mind that allowing for heterogeneous treatment effects by including the supply of teachers may introduce a problem related to unobserved sorting of teachers within school districts if the sorting depends on the chosen hiring practice. This may obviously have some implications for the analysis, and will be discussed in more detail in the empirical results in Section 5.

5. Empirical Results

5.1 School district level analysis

Table 2 reports the results from estimating various specifications of equation (1) using district level data. In column (I) school district efficiency is regressed only on the dummy for decentralized hiring. The difference in efficiency between districts with decentralized and centralized hiring practices is estimated to be 9.5 percentage points corresponding to the raw difference in the data (see Figure 1b). As reported in Figures 1a and 1b, the decision to decentralize is positively correlated with both educational efficiency and school district size, and the estimated effect of decentralized hiring in column (I) may therefore capture the effect on efficiency of school district size. In column (II) I therefore include the (log) number of students as a control for school district size. The estimated effect of school district size is positive and highly significant. The estimated effect of decentralized hiring is reduced by more than 40 per cent (four percentage points) when school district size is included. In column (III) the squared log of the number of students is added. The quadratic term has a positive sign and is significant at the 10 per cent level, providing some support for a non-linear relationship. The estimated effect of hiring practices is only marginally changed compared with the effect in column (II). [13] As the support for a non-linear relationship is stronger when additional controls are included, I choose to continue with this specification of the model.

Next, in column (IV) I include the share of students with special needs and the share of people with higher education in the school district. The information about students with special needs is only provided at the aggregate level (for schools and school districts). Special needs therefore cannot be included for individual students separately in the achievement regressions prior to the DEA analysis. As in Borge and Naper (2006), the estimated effect is negative and statistically significant. A high share of students with special needs is associated with lower school district efficiency. The level of education in the school district is, as discussed, an important control variable if the education orientation of the electorate is reflected by the organization of schools. However, the level of education is found to have an insignificant effect on the efficiency level and the estimated effect of decentralized hiring is only marginally reduced compared with the results in column (III).

In column (V) I include some additional controls. These are the share of people in the school district living in rural areas (settlement pattern), political variables (party fragmentation and socialist share), and a measure of local revenues. The reported results concerning these variables are comparable with the findings presented by Borge and Naper (2006). The share of people living in rural areas has a significant negative effect. Local revenues are included as income taxes and block grants per capita deflated by an index of varying cost conditions across local authorities. The estimated effect is negative and statistically significant. Hence, school districts with a high level of revenues tend to be less efficient. The underlying argument for this relationship may be that the schools (service producers) are able to take advantage of a rich sponsor to increase budgetary slacks. The positive sign of the Herfindahl index of party fragmentation means that school districts with a highly fragmented local council tend to have low efficiency. The positive sign of the socialist share indicates that socialist influence is associated with low efficiency. However, the effects of political variables are only marginally significant.

[13] Including higher powers of the log number of students does not change the estimated coefficient.

Table 2
The effect of decentralized hiring on school district efficiency

	(I)	(II)	(III)	(IV)	(V)	(VI)
Decentralized hiring practices	0.095	0.054	0.051	0.047	0.038	0.038
	(6.02)	(3.68)	(3.48)	(3.20)	(2.79)	(2.63)
Log of number of students		0.068	–0.046	–0.046	–0.213	–0.303
		(8.44)	(0.67)	(0.68)	(2.86)	(2.53)
Squared log of number of students			0.009	0.008	0.019	0.028
			(1.67)	(1.47)	(3.43)	(2.78)
Share of population with higher education				0.177		
				(0.90)		
Share of students with special needs				–0.749	–0.576	–0.564
				(3.03)	(2.46)	(2.37)
Local revenues					–0.225	–0.228
					(5.72)	(5.64)
Share of population living in rural areas					–0.077	–0.074
					(2.47)	(2.34)
Socialist share					–0.074	–0.076
					(1.57)	(1.57)
Herfindahl index of (inverse) party fragmentation					0.136	0.129
					(1.58)	(1.47)
Log likelihood	151.64	183.56	186.55	190.03	204.83	195.71
Observations	250	250	250	250	248	238

Note: Tobit estimates with absolute t-values in parenthesis. Column (VI): estimates based on a reduced sample excluding 10 school districts.

The estimated effect of hiring practices reported in column (V) is around 1 percentage points lower compared with the effect in column (IV), and the results suggest that school districts with decentralized hiring practices on average are 3.8 percentage points more efficient than school districts with traditional centralized hiring practices. The estimated effect is statistically significant at the 1 per cent level, and amounts to approximately 40 per cent of the observed raw discrepancy in efficiency.

A close look at the data reveals that among the school districts in the upper part of the district size distribution there are no districts with a centralized hiring practice in the estimation sample. These districts are without an obvious comparison group in the estimation sample. Although the columns (II)–(V) control for school district size, one might still be concerned that the estimated effect of decentralized hiring is driven by the largest districts, where only decentralized hiring practices are observed. To investigate whether the estimated effect in column (V) is influenced by this lack of common support in the data, I therefore re-estimate the model in column (VI) using a reduced sample. The sample is reduced by excluding the largest school districts with decentralized hiring practices. In total, 10 school districts are excluded from the sample and the last school district excluded has 42 745 inhabitants (2 300 students). The effect of hiring practices in column (VI), where 10 school districts have been excluded, is only slightly lower compared with the effect in column (V) and is still significant. The effect of decentralized hiring does not seem to be driven by the predominance of larger school districts that decentralize in the sample.

The empirical findings in Table 2 provide support for the hypothesis that school districts that choose decentralized hiring practices on average are more efficient than school districts with centralized hiring practices. Furthermore, the results are consistent with the theoretical predictions in Bishop and Wößmann (2004), and empirical contributions by Robin and Sprietsma (2003) and Zigarelli (1996).

5.2 School level analysis

As already emphasized the Tobit estimates in Table 2 may be biased if the decision to decentralize is based on unobservable school district characteristics that are correlated with efficiency. This potential problem is addressed by allowing for heterogeneous treatment effects in an individual school level analysis as proposed in Section 4. Schools within the same school district cannot generally be treated as independent observations, and the model should be estimated with cluster-robust standard errors. To take into account that the data is both censored and clustered, the preferred estimation strategy is a Tobit model with clustering. This is, however, not easily calculated and the alternative strategies are to use either OLS with clustered standard errors or Tobit without clustered standard errors. OLS estimation in the case of a censored dependent variable generally produces a downward bias in the estimates, while estimation without cluster-robust standard errors when the estimation data are multilevel typically produces a downward bias in the standard errors. As the share of censored observations is relatively small (1.2%) I have chosen to rely on OLS with clustered standard errors. Table 3 presents the results from estimating equations (2) and (4) by OLS with cluster-robust standard errors. However, a specification of the model based on Tobit is also reported in column (III) of Table 3, indicating that the OLS estimation bias is limited.

First, the model is estimated as described in equation (2) without district FE. The dependent variable used in Table 3 is the efficiency scores for individual schools and the same set of school district characteristics used above is included. In column (I) I simply reproduce the school district estimation reported in column (VI) of Table 2 using school level data. The effect of decentralized hiring is 2.1 percentage points and marginally significant. Compared with Table 2, the qualitative effect is reduced by almost 40 per cent. The number of students and the share of students with special needs are included separately for individual schools and not as school district means. In addition, I include the educational level in each school's catchment area.[14] The education orientation in the population may vary within school districts as well as across districts, and the variable is included because the education orientation of the school users may affect the distribution of teachers across schools as well as how school leaders choose to manage their schools. The educational level of the school catchment area has a positive but insignificant effect. The remaining control variables in column (I) have comparable effects as in Table 2.

In column (II) I allow for heterogeneous treatment effects by including the dummy variable for excess teacher supply presented in Figure 2. The variable is included separately and interacted with the dummy for decentralized hiring. The effect of the dummy variable for decentralized hiring reported in column (II) should be interpreted as the effect of decentralized hiring for schools with excess teacher *demand*. The estimated effect is close to zero and insignificant as expected. The effect of decentralized hiring for schools with excess teacher supply is found by adding the estimated interaction between excess supply and decentralized hiring practices. The interaction effect is positive and statistically significant at the 5 per cent level, and indicates that the effect of decentralized hiring on efficiency is stronger when schools can choose between several qualified applicants.

Column (IV) in Table 3 presents the results from estimating equation (4) where district FE are included to exploit within school districts variation only. All school districts with only one school automatically drop out of the analysis as there is no within school district variation. The number of observations therefore drops slightly compared with column (II). The estimated interaction effect comes out larger and highly significant in the FE specification. Hence, when unobserved school district characteristics and potential sorting of teachers across school districts is controlled for, the effect of decentralized hiring is found to depend positively on teacher supply. The estimated interaction effect indicates that the positive effect of decentralized hiring is 7.5 percentage points larger for schools with excess teacher supply than for schools with excess demand.

[14] The educational level in school catchment areas is measured by the share of students at each school that have parents (mother, father, or both) with education beyond the upper secondary school level.

Table 3
The interaction effect between decentralized hiring practices and teacher supply

	(I) OLS	(II) OLS	(III) Tobit	(IV) FE	(V) FE	(VI) FE	(VII) FE
Decentralized hiring practices	0.021	0.008	0.008				
	(1.63)	(0.50)	(0.59)				
Interaction term: decentralized hiring practices and excess teacher supply		0.043	0.044	0.075	0.081	0.060	0.063
		(2.11)	(2.33)	(3.22)	(2.96)	(2.41)	(2.03)
Dummy for excess teacher supply		–0.003	–0.003	–0.034	–0.042	–0.032	–0.018
		(0.18)	(0.19)	(1.88)	(1.85)	(1.83)	(0.72)
Log of number of students	0.415	–0.413	–0.416	–0.277	–0.326	–0.375	–0.281
	(5.85)	(6.03)	(8.99)	(3.67)	(4.21)	(4.70)	(3.55)
Squared log of number of students	0.046	0.046	0.046	0.029	0.033	0.039	0.029
	(6.33)	(6.53)	(9.17)	(3.71)	(3.99)	(4.39)	(3.58)
Share of population living in rural areas	–0.102	–0.097	–0.096				
	(4.15)	(3.89)	(3.82)				
Herfindahl index of (inverse) party fragmentation	0.105	0.118	0.116				
	(1.43)	(1.61)	(1.51)				
Socialist share	–0.000	0.001	0.001				
	(0.00)	(0.01)	(0.04)				
Municipal income	–0.173	–0.171	0.170				
	(5.13)	(5.47)	(4.92)				
Share of population with higher education (school district)	0.019	0.035	0.039				
	(0.14)	(0.29)	(0.40)				
Mean share of population with higher education (school's catchment area)	0.131	0.123	0.123	0.172	0.111	0.086	0.169
	(1.42)	(1.38)	(3.40)	(1.78)	(1.24)	(1.38)	(1.69)
The share of students with special needs	–0.472	–0.450	–0.449	–0.842	–0.764	–0.646	–0.809
	(3.63)	(3.48)	(3.85)	(4.80)	(3.80)	(2.70)	(4.57)
Mean age of teachers					0.003		
					(1.65)		
The share of teachers with 4 years' higher education					0.173		
					(2.08)		
The share of teachers with 5 years' higher education					0.129		
					(1.94)		
The share of teachers with 6 years' higher education					0.521		
					(5.63)		
School district fixed effects	NO	NO	NO	YES	YES	YES	YES
Observations	653	650	650	552	467	396	549
R-squared	0.42	0.43	–	0.20	0.23	0.21	0.20
Log likelihood	–	–	491.49	–	–	–	–

Note: Estimates with absolute t-values in parenthesis. Standard errors are heteroscedasticity robust and adjusted for school district level clustering. The models include a constant term. In column (V) there are fewer observations because not all schools can be matched with teacher information. Column (VI) is estimated using the reduced sample as in Table 2. In column (VII) the model is estimated using an extended indicator for excess teacher supply.

Norwegian teachers are generally linked to only one school and normally cannot be instructed to move from one school to another. Teacher transitions within school districts are therefore, with very few exceptions, voluntary and initiated by the teacher. Important determinants for within school district mobility of teachers are student composition and student achievement (Hanushek et al., 1999). The models presented in columns (I)–(IV) do not consider teacher mobility, and it could be argued that this would bias the estimates. If more efficient schools attract more qualified teachers it is quite obvious that the estimated dummy for excess teacher supply will be biased upwards. However, in relation to hiring practices, it is the interaction term between decentralized hiring and excess supply that is the main interest. It is less clear that this coefficient would be biased because of teacher mobility. The coefficient would only be biased if the sorting of teachers differs substantially between districts with different hiring practices.[15] It would be upwardly biased if the tendency for efficient schools to attract more qualified teachers is stronger in districts with decentralized hiring practices.

If the interaction effect is estimated with an upward bias in column (IV) we should expect that the effect would decrease if teacher quality was controlled for. Teacher quality is generally hard to measure perfectly. However, in column (V) I include a set of observable teacher characteristics which are likely to be correlated with teacher quality. These are the average age of the teaching staff and the average educational length of the teachers. Educational length is measured by the share of teachers with a regular four year Teachers' college degree, the share of teachers with additional training after Teachers' college (five years), and the share of teachers with a higher university degree (six years or more). The average age of the teaching staff has a positive sign but is only marginally significant. Teachers' educational length is estimated to have a significant and positive effect on educational efficiency. Importantly, the estimated interaction effect is slightly higher and still statistically significant when observable teacher quality is included.

5.3 Robustness of the interaction with teacher supply

Two additional robustness checks of the estimated interaction effect are presented in Table 3. First, decentralization is widespread in the largest school districts, and teacher supply is also slightly higher than overall in schools within these districts. Although the results in Table 2 indicate that the effect of hiring practices is not driven by the largest school districts in the sample, the estimated interaction with teacher supply may be partly related to schools in the largest districts. An important robustness check is therefore to exclude the largest school

[15] This can, for example, be the case if school districts with traditional centralized hiring practices are, to a larger extent than school districts with decentralized practices, able to affect the allocation of teachers between schools by sorting the best teachers to the schools with low-performing students.

districts from the sample. I therefore re-estimate the model using the reduced sample as was done earlier. The number of schools in the estimation sample is reduced by almost 30 per cent as the larger school districts have more schools. The estimated interaction effect is reported in column (V) and is slightly lower but still significant at the 5 per cent level.

Second, decentralization is expected to work better for schools with excess teacher supply, and the effect should therefore be stronger the higher the persistence in teacher supply. Therefore, there is an argument to include teacher supply further back in time to obtain a more long-run measure of persistence. In column (VI) I therefore re-estimate the model using a dummy variable for excess teacher supply indicating whether the schools had excess teacher supply in the final three years instead of the final two years. The interaction effect is slightly lower compared with column (IV), but is still statistically significant.

6. Concluding Remarks

Numerous papers have analysed the importance of teacher qualifications for student performance and educational efficiency. A relatively large literature also focuses on the hiring practice of teachers and how to recruit effective teachers in public schools. However, relatively few papers have studied how the design of the teacher hiring process affects educational efficiency. This paper contributes to the literature by analysing how decentralized hiring practices of teachers affect educational efficiency in Norway. Norwegian school districts are free to delegate responsibilities of school management decisions to individual schools and decentralized hiring practices mean that the hiring decisions are taken at the school level by the principal. School principals may know better than the school district administration which qualifications are most required for a given teaching position and may therefore more easily evaluate the applicants for a given position. Decentralized hiring practices are then expected to positively affect efficiency because of an overall better match between schools, teachers, and students.

Because decentralization is a local choice and information about the factor(s) that triggers the decision is unavailable, the causal effect of hiring practices on educational efficiency may be hard to pin down and naïve estimates may suffer from an endogeneity bias. In the empirical analysis I first address this problem by including a rich set of observable school district characteristics in a school district analysis. The estimation results clearly show that school districts with decentralized hiring practices on average are more efficient than districts with traditional centralized hiring practices. Furthermore, I realize that if the decision to decentralize is based on unobservable school districts characteristics that are correlated with educational efficiency, the estimates may still be biased. In a second approach I therefore allow for heterogeneous treatment effects for individual schools in a school level analysis where I exploit the fact that the impact of decentralized hiring practices is expected to be stronger when schools have excess teacher supply. Including school district fixed effects I can estimate the interaction effect between excess supply and hiring practices, although not the direct effect of hiring practices, using only within school district variation. The empirical results support the hypothesis that decentralized hiring practices have a stronger impact when schools can choose between several qualified applicants.

References

Ballou, D. (1996): Do Public Schools Hire the Best Applicants?, The Quarterly Journal of Economics, 111 (1), pp. 97-133.

Balter, D. and Duncombe, W. (2006): Staffing Classrooms: Do Teacher Hiring Practice Affect Teacher Qualifications?, Working Paper, American Education Finance Association.

Bettinger, E. (2005): The Effect of Charter Schools on Charter Students and Public Schools, Economic of Education Review, 24(2), pp. 133-147.

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

Bonesrønning, H., Falch, T. and Strøm, B. (2005): Teacher Sorting, Teacher Quality and Students Composition, European Economic Review, 49, pp. 457-483.

Borge, L. E. and Naper, L. R. (2006): Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools, FinanzArchiv, 62 (2), pp. 221-249.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978): Measuring the Efficiency of Decision Making Units, European Journal of Operational Research 2, pp. 429-444.

Christiansen, T. and Lægreid, P. (2005): New Public Management, Ashgate Publishing Limited.

Clark, D. (2005): Politics, Markets and Schools: Quasi-Experimental Evidence on the Impact of Autonomy and Competition from a Truly Revolutionary UK Reform, WP Nuffield College, Oxford.

Education at a Glance Indicators (1998): Report, Organization for Economic Co-operation and Development.

Eskeland, G. S. and Filmer, D. (2007): Autonomy, Participation and Learning: Findings from Argentine Schools, and implications for decentralization. Education Economics, 15 (1), pp. 103-127.

Haraldsvik, M. and Naper, L. R. (2007): Effektivitet og Nye Organisatoriske grep, Center for Economic Research, report 02/07.

Hanushek, E. A., Kain, J. F., Rivkin, S. G. and Branch, G. F. (2007): Charter School Quality and Parental Decision Making with School Choice, Journal of Public Economics, forthcoming.

Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (1999): Do Higher Salaries Buy Better Teachers? , Working Paper 7082, National Bureau of Economic Research.

Hoxby, C. M., and Rockoff J. E., (2004): The impact of Charter Schools on Students Achievement, Working paper Department of economics, Harvard University, Cambridge.

Lyng, S. T. and Blichfeldt J. F., (2003): Attracting, Developing and Retaining Effective Teachers: Country Background Report for Norway, Organization for Economic Co-operation and Development March 2003.

Rivkin, S. G., Hanushek, E. and Kain, J. F. (2005): Teachers, Schools and Academic Achievement, 73 (2), pp. 417-458.

Robin, R. R. and Sprietsma, M. (2003): Characteristics of Teaching Institutions and Students' Performance: New Empirical Evidence from OECD Data, Université Catholique de Lovain, Institute de Recherches Economiques et Social (IRES), mimeo.

Rockoff, J. E., (2004): The Impact of Individual Teachers on Students Achievement: Evidence from Panel Data, American Economic Association, papers and proceedings, 94 (2).

Worthington, A. C. (2001): An Empirical Survey of Frontier Efficiency Measurement Techniques in Education, Education Economics 9, pp. 245-268.

Wößmann, L. (2003): Schooling Resources, Educational Institutions and Student Performance: The International Evidence, Oxford Bulletin of Economics and Statistics, 65 (2), pp. 117-170.

Zigarelli, M. (1996): "An empirical Test of Conclusions From Effective School Research", The Journal of Educational Research, 90 (2), pp. 103-110.

Appendix

Table A1

*Summary statistic school and school district characteristics**

Variable	Decentralized	Centralized	Not in survey
School district level			
Number of school districts	113	137	181
Population size (in 000s)	21.89 (2.59)	5.28 (1.08)	7.52 (1.38)
Number of students in lower secondary schools (school district mean)	1052 (2.22)	335 (0.80)	440 (1.06)
Efficiency score (DEA)	0.76 (0.16)	0.67 (0.18)	0.69 (0.18)
The share of population living in rural area	0.39 (0.67)	0.54 (0.48)	0.51 (0.53)
Municipal income (mean 100)	1.02 (0.18)	1.11 (0.19)	1.07 (0.15)
The share of representatives from the socialist camp in local council (based on the 1999 local election)	0.36 (0.34)	0.37 (0.42)	0.36 (0.39)
The (inverse) Herfindahl index of party fragmentation in the local council (based on the 1999 local election)	0.24 (0.41)	0.29 (0.40)	0.26 (0.40)
Share of students with special needs (school district mean)	0.06 (0.52)	0.07 (0.43)	0.07 (0.43)
Number of teacher hours per student	78.38 (0.18)	87.03 (0.19)	86.01 (0.19)
The share of population with higher education	0.18 (0.31)	0.15 (0.24)	0.17 (0.30)
School level			
Number of schools	391	262	408
Number of students	325 (0.34)	206 (0.47)	236 (0.46)
The share of students with special needs	0.07 (0.57)	0.08 (0.56)	0.07 (0.90)
The share of certified teachers in individual schools	0.98 (0.02)	0.97 (0.04)	0.96 (0.05)
Number of teacher hours per student	69.35 (0.15)	87.79 (0.19)	84.79 (0.25)
The share of students with highly educated parents at the school (school catchment area)	0.39 (0.40)	0.30 (0.45)	0.31 (0.47)
The average age of the teachers	45.78 (0.07)	45.61 (0.10)	45.32 (0.16)
The share of teachers with 4 years' higher education	0.38 (0.34)	0.40 (0.43)	0.41 (0.70)
The share of teachers with 5 years' higher education	0.36 (0.36)	0.31 (0.61)	0.31 (0.90)
The share of teachers with 6 years' higher education	0.07 (1.00)	0.04 (1.75)	0.04 (2.75)
Efficiency scores (DEA)	0.69 (0.21)	0.59 (0.22)	0.62 (0.21)

* Mean values with coefficient of correlation in parenthesis. The DEA efficiency scores are based on data from the academic years 2002–2003 and 2003–2004. All demographic and economic characteristics used in the empirical analysis are averaged over the same years.

Chapter 4
Heterogeneous Effects from School Competition

Heterogeneous Effects from School Competition

Hans Bonesrønning* **and Linn Renée Naper**

Abstract

We examine whether potential competition effects from private schools differ across student subgroups in Norwegian public schools. The endogeneity of competition - measured by the share of students attending private schools – is treated by IV estimation. Sorting of parents across municipalities is considered by exploiting the location of universities and colleges. Indications of non-uniform competition effects are provided: while there is no evidence that students with less educated parents gain from competition, the effect for students with highly educated parents is positive. The robustness analyses indicate that the latter effect to some extent depends on the outcome variable used.

Keywords: school competition; private schools; student achievement.
JEL-classifications: I21, H42

1. Introduction

In many countries school reformers introduce competition from charter schools as a means to improve the performance of public schools. However, it is a controversial issue whether school competition will work as intended. From a theoretical point of view, the nature of competition between schools is not clear-cut: The education market seems to be a differentiated product market, where schools differ in their characteristics and where parental preferences over school characteristics differ. One conjecture is therefore that the degree and character of competition will vary with the characteristics of both suppliers and demanders. Not surprisingly then, the existing empirical literature on school competition – between public and private schools or between public schools – does not provide any sweeping and clear-cut conclusions (see for instance, Hoxby 1994, 2000; Dee 1998; Fiske and Ladd 2001; Levačić 2004; Bettinger 2005; Sandstrøm and Bergstrøm, 2005; Bayer and McMillan 2005, Gibbons, Machin and Silva, 2005; Hsieh and Urquiola, 2006; Card, Dooley and Payne, 2006).

* An earlier version of this paper has been presented at the 2nd Workshop of the RTN 'Education Economics and Education Policy in Europe' and at Tinbergen Institute in Amsterdam. We thank the participants for insightful comments. Any errors are ours.

The present paper adds to the existing literature by providing empirical evidence from Norway. First, we investigate whether the overall performance of public schools in Norway is affected by competition from private schools. Thereafter, we ask whether the competition effects differ across student subgroups. Our hypothesis is that principals and teachers within public schools might realize that they can influence parental school choices by reallocating resources and/or attention towards the students they think are most likely to opt out. The incentives to reallocate resources or attention might vary with the attractiveness of the most mobile students (it is not unlikely that mobility and attractiveness is correlated). Thus, if the public sector actors' responses to competition are to reallocate resources between student subgroups, we expect the attractive and mobile student subgroups to gain, and the less attractive and less mobile students, to lose from competition.

The private school sector in Norway is small, and has some distinct characteristics. Most privately owned schools are religious schools or schools with an alternative pedagogy, reflecting the Private School Law of 1970 which offers only these kinds of private schools public funding. Private schools are offered (almost) free of charge to the students. The municipalities experience a decrease in their budgets when students exit for private schools; implying that the financing of private schools very much portrays a voucher system. Private schools are not allowed to discriminate between applicants, but the rules that can be applied when schools are oversubscribed are vague.

At first sight Norwegian private schools do not look like close substitutes to public schools, and one might conjecture that the private schools are not much of disciplinary devices for the actors in public schools. A simple argument is that "religious parents prefer religious schools". End of story. However, for marginal students at least, the school choice is likely to involve trade-offs between the school's programme, student achievement, travelling distances and so on. Whether there are any competition effects is an empirical issue.

We argue that the local markets are defined by municipality borders, and use the share of students attending private schools in the municipality as our measure of competition. A major econometric challenge is that private schools are not randomly distributed across the municipalities. We show that municipalities with private schools differ from municipalities without private schools in many observable dimensions. In addition, we suspect that the private schools are established for one of the following reasons: i) poor public school quality (that is hardly observable to the researcher), or ii) high demand for educational quality due to unobservable parental preferences.

This is to say that the measure of competition is potentially endogenous to student achievement in public schools. As our point of departure, we follow the Swedish study by Sandstrøm and Bergstrøm (2005), and utilize the contracting-out of child care services as an instrument[1]. Their argument is that although the establishment of Swedish independent schools is determined by national authorities, the local authorities are allowed to give their opinion on whether they consider the establishment of an independent school to be harmful to existing schools. Also, there are informal ways in which a municipality can aid or hinder the establishment of a private school, e.g. by delaying the necessary permits for the use of buildings by a school. The establishment of private schools is thus facilitated in municipalities where local authorities have positive attitudes towards such schools (indicated by the contracting-out of child care). The national and local authorities play the same roles in Norway as in Sweden. The exclusion restriction is that the share of children in private child care does not exercise any direct influences on student achievement in the 10^{th} grade.

In addition to the contracting-out of child care we introduce three more identifying variables. First, we argue that the share of representatives from the right wing party "Fremskrittspartiet" (FRP) in Norwegian local councils may serve as an identifying variable. FRP is a right-wing, populist party, that strongly support the establishment of private schools. The party's support for private schools seems to be driven by ideology and not by their electorate's education-orientation. Second, we follow a number of US studies that have used a set of denominational variables as instruments to deal with the endogeneity problem. The motivation is that many private schools in the US are run by Catholic institutions. Since as many as sixty of the Norwegian private schools are run by religious institutions, we propose the share of representatives from the Christian Democratic Party in the local council as an identifying variable. Finally, we exploit that Norway is a scattered populated country. Private schools may be hard to establish in municipalities with long travelling distances. A reasonable assumption is that the local demand for private schools shifts inwards as the travelling distances within the municipality increase, thus generating the necessary correlation between the share of students in private schools and travelling distances. The instruments – particularly the exclusion restrictions - are discussed more thoroughly in a subsequent section.

[1]Actually, they propose five indicators of the municipalities' willingness to contract out their responsibilities, but only the contracting-out of child care is significantly associated with the share of students in independent schools.

There are more, and related, problems to be solved; first and foremost related to the share of highly educated citizens in the municipality. As will be documented later, this variable is correlated both with our measure of competition and with student achievement in public schools, implying that the estimated competition effects are not robust to specification. The latter of these correlations might reflect that parents and/or teachers have sorted themselves across municipalities due to unobservable preferences for public school quality, while the former correlation might reflect that the sorting is imperfect. That is, some inhabitants may have responded to poor public school quality by establishing private schools, and not by moving out of the municipality. The share of highly educated citizens in the municipality is then endogenous to student achievement in public schools. Introducing this variable in the estimated equation will not work to generate an unbiased estimate of the competition effect. We suggest using the location of universities and colleges to get around this endogeneity problem. In addition, we report the results from equations with and without the share of highly educated citizens among the explanatory variables. In the former specification, the competition effects most likely are biased downwards, in the latter they are most likely biased upwards.

We find that in no specifications are the competition effects for students with less educated parents positive. For students with highly educated parents, the competition effects are always positive. In some specifications these latter effects are small and insignificant; in others they are quite large and significant. We also show that students with highly educated parents are among the students that have the highest propensities of leaving public schools. Taken together, these findings lend support to our hypothesis that public schools respond to competition by reallocating resources towards the most attractive and mobile students.

The remainder of this paper proceeds as follows. Chapter 2 gives a short overview of the Norwegian school sector and presents our data. Chapter 3 outlines our empirical strategy and discusses econometric challenges. Chapter 4 presents the empirical results. Chapter 5 presents robustness checks of the empirical results and section 6 offer some final conclusions.

2. Private and Public Schools in Norway

2.1 Institutional details

Most primary (1.-7. grade) and lower secondary schools (8.-10. grade) in Norway are owned by the municipalities. The municipalities decide the number of schools, school location and school budgets, but the workings of the schools are heavily influenced by the national government: The multi-purpose municipalities' revenues are almost fully determined by grants and local tax rates set by the Government. Moreover, the public schools operate under strict national rules and regulations that strongly influence the number of students per teacher, teacher pay and teacher qualifications. The curriculum is decided at the national level, and there are national exams in mathematics, reading and writing and English language at the end of the 10th grade. A neighbourhood school rule has been applied throughout the country for the actual period of time, making many public schools local monopolies[2].

The private school law of 1970 restricted the public financing of private schools to religious schools or schools with a distinct pedagogical idea (such as Steiner and Montessori)[3]. When approved by the Ministry, private schools receive public funding amounting to 85% of the municipality's costs for schooling per student. The financing system encourages the establishment of small private schools by compensating for diseconomies of scale. The private schools thus receive more funding per capita for the first 40 students than they do for the next 160 students. The schools are allowed to charge a small tuition fee, but the sum of public funding and tuition fees cannot exceed the costs per student in comparable public schools in the municipality. The private schools are financed by the Government, but the municipalities experience a decrease (less than the average cost per student) in their grants when students exit for private schools.

The approval of new private schools is made by the Ministry of Education, but the municipalities are allowed to give an opinion on whether they consider the establishment of a private school to be harmful for existing schools (The Private School Law §25). The municipalities have no veto, and the Ministry can choose to follow the will of the municipalities or not.

2.2 Private schools' location

The Norwegian primary and lower secondary schools have compiled thorough school level statistics since the academic year 1992-1993 only. Table 2.1 shows that the number of private schools have more than doubled from 1992-1993 to 2003-2004, from 49 to 116 schools. By 2003-2004 a little more than 50 percent of the private schools were religious schools, and a

[2] In comparison, Sweden introduced a parental choice reform in the early 1990s.
[3] The present study applies data for the school year 2003/2004, which is the last year before a new and more liberal private school law was introduced.

little more than 50 percent of the students in private schools were enrolled in religious schools.

Table2.1
The number of privately operated schools from 1992-1993 until 2003-2004

Academic year	Number of privately operated schools (primary and lower secondary).	Number of privately operated schools with students on the lower secondary level (grade 8 to 10)
1992 – 1993	49	34
1993 – 1994	55	41
1994 – 1995	63	48
1995 – 1996	65	48
1996 – 1997	67	51
1997 – 1998	68	52
1998 – 1999	71	57
1999 – 2000	81	61
2000 – 2001	89	65
2001 – 2002	98	68
2002 – 2003	106	71
2003 – 2004	116	83

A first view on private school location is offered in Table 2.2. The private schools are concentrated in 76 (from a total of 434) municipalities. For municipalities that have private schools, the private market share is 4.7 percent on average, varying from below 1 percent of the total student population in the municipality up to nearly 19 percent. The private schools tend to be located in large municipalities, as indicated by the fact that almost half the student population is located in the municipalities with private schools, and 25 percent of the municipalities with private schools having more than 30 000 inhabitants. It cannot be read from the table that a fairly large number of private schools are located in small municipalities in remote areas, but from the raw data we see that 25 percent of the municipalities with private school alternatives have a population of less than 7000 inhabitants.

Table 2.2 also shows that the population compositions differ in potentially important respects between municipalities with and without private schools. Municipalities with private schools have a more highly educated population than municipalities with only public schools, while the dispersion in people's educational level seems to be somewhat higher in municipalities with no private schools. The share of immigrants turns out to be significantly larger in the municipalities with private schools than in municipalities without private schools.

Table 2.2
Summary statistics for municipalities

	Municipalities with no independent schools					Municipalities with one or more independent schools				
Total number of municipalities:	359					74				
Total number of 10th graders :	25 571					26 524				
Variable	# of valid obs.	Mean	Standard deviation	Min	Max	# of valid obs.	Mean	Standard deviation	Min	Max
The share of students in independent schools	359	0	0	0	0	75	0,047	0,036	0,002	0,185
Market share for private kinder gardens (contracting out)	359	0,194	0,227	0,000	1,000	75	0,418	0,198	0,000	0,848
The share of representatives in local council from the right wing party FRP	356	0,054	0,066	0,000	0,400	74	0,118	0,060	0,000	0,259
The share of representatives in local council from the Christian Democratic Party	356	0,093	0,092	0,000	0,560	74	0,098	0,073	0,000	0,379
The share of socialists in the local council	356	0,364	0,146	0,000	0,846	74	0,374	0,118	0,135	0,622
Travelling distance within zone per citizens	358	8,836	7,246	0,319	59,873	75	4,864	4,175	0,987	20,563
Herfindahl Index of public school competitiveness in municipality	359	0,411	0,259	0,050	1,000	75	0,169	0,156	0,009	1,000
The share of employees between 20 and 54 in the municipality	359	0,615	0,045	0,438	0,778	75	0,624	0,046	0,400	0,687
The share of population with higher education	359	0,187	0,042	0,091	0,390	75	0,240	0,074	0,113	0,467
Herfindahl Index of educational level in the municipality	357	0,161	0,035	0,108	0,500	75	0,135	0,022	0,094	0,197
The share of immigrants in the municipality	359	0,033	0,019	0,003	0,159	75	0,050	0,030	0,007	0,197
The school cost per student in the municipality	355	66,481	14,252	46,971	130,244	75	57,542	10,719	44,743	109,641
Municipal revenue	359	107,499	17,615	92,000	214,000	75	99,147	8,859	93,000	129,000
Population size (in 1000)	359	5,857	7,025	0,215	56,688	75	32,998	67,745	0,887	521,886
The number of students in municipality	359	824,304	974,891	32	6661	75	4122,08	7014,969	126	49127

Finally, note that Table 2.2 shows that municipalities with private schools on average have larger shares of private kinder gardens, larger shares of representatives from the right wing party FRP in the local councils, and shorter travelling distances between local centres. These are very useful features for the econometric analyses. Newmark (1995) argues that competition is unlikely to have any noticeable effect when enrolment in private school does not vary over time, and moreover, that studies of the potential private competition effect on public schools are unlikely to find any significant effect due to low variability in the data. The first view on the Norwegian data shows that there has been quite a rapid growth in the number of private schools in the years prior to 2003/2004; which is the year we analyze, and that there is quite a substantial cross section variation in the share of students in private schools across municipalities.

2.3 Public school performance

The performance of public schools is measured by student achievement at the end of the lower secondary school. We have several possible achievement measures to use. For the school year 2003/2004 we have for the 10th graders access to assessment grades (grades set by the teachers) in all subjects on the curriculum, external examination results in core subjects and results from a national test in reading and writing, mathematics and English, all at the individual level.

The natioanl tests are preferred in the analyses, because they are more objective than the teachers' grades (see Wikstrøm and Wikstrøm (2005) for Swedish evidence about grade inflation and school competition), and because they cover larger fractions of the students than do the exam results (for a considerably large part of the municipalities only a subset of the subjects are covered by the national exam). However, not all students have valid test results. Therefore we use the students' grades to evaluate the robustness of our findings (moreover, the grades are compared to the students' exam results to evaluate the responsiveness of the teachers' grading practices to competition pressures).

The student level data file contains individual test scores from national achievement tests, covering a total of 51 322 10th graders from public schools. 43 113 students have valid test results in mathematics, 44 654 in reading and writing and 44 119 in English. The number of students that have valid test results in all three subjects is 34 035. Appendix Tables A.1-A.3 show mean achievement levels in municipalities with different fractions of private school enrolments.

2.4 Other variables

The test results are linked to data from Statistics Norway on individuals and families. We have information about the students' sex, date of birth, number of siblings, birth order, and their migration status, and further, the parents' educational level and income (separate for mother and father), marital status, number of working hours per week, which sector they work in (private/public) and whether they are self-employed or not.

Data describing school inputs come from The Ministry's data base for the elementary school (GSI). The school level data file contains both public and private schools and provides information about the number of students, the share of students with immigration status (students receiving separate teaching in Norwegian language) and the share of students that receive additional resources due to learning disabilities, the number of teachers, and the number of certified teachers. The school level data cover around 1100 primary and lower secondary schools. Data describing the municipalities come from the Norwegian Social Sciences Data Services (NSD). The municipalities' characteristics reported in table 2.2 above come from this data source. Moreover, we have access to data of travelling distances between schools within municipalities. The municipal level data cover all the 434 municipalities in Norway.

3. Empirical Strategy

We follow Sandstrøm and Bergstrøm (2005) and let the share of students attending private schools in each municipality be a gauge of the degree of competition. As in Sweden, very few Norwegian students go to schools outside the municipality where they reside, clearly making the municipality the relevant market. The measure of the private market share in our analyses is based on all students attending compulsory schools (1^{st} to 10^{th} grade) in the respective municipalities. We investigate how the variation in private sector enrolment affects the educational outcomes in public schools in the following individual level regression model:

$$(1) \qquad t_{ism} = \alpha p_m + \beta_1 B_{ism} + \beta_2 S_{sm} + \beta_3 M_m + \varepsilon_{ism}$$

Equation (1) is a variant of the well-developed and broadly discussed Education Production Function (see for instance Hanushek, 2002) where schools are regarded as firms producing educational outcomes in terms of student achievement. The variable t_{ism} is student i's test score at the public school s in municipality m. The variable p_m is the private market share in the municipality m where student i goes to school. The three vectors B, S and M include family background variables, school variables and municipality variables respectively. The α and $\beta's$ are parameters to be estimated and ε_{ism} is a composite error term.

As stated above, the share of students in private schools is potentially an endogenous variable in equation (1). We construct an instrumental variable by estimating an equation that seeks to explain the variation in private market shares across municipalities in the following TOBIT specification[4]:

$$(2) \qquad p_m = \delta D_m + \nu_m$$

The variable p_m is the private market share used in equation (1) which cannot be lower than zero and thus is censored at 0; D is a vector of potential determinants for private schools establishment and ν_m represents a white noise error term.

To solve the endogeneity problem, the D-vector must contain at least one variable that is correlated with the private market share and fulfils the exclusion restriction. As stated above, we exploit that local politicians have a legal right to influence the establishment of private schools, and seek to characterize local politicians in ways that are not correlated with student achievement. Initially, we follow Sandstrøm and Bergstrøm (2005) who claim that the extent to which the municipalities contract out their responsibilities indicate their attitude to the privatization of public sector activities. Contracting out in other sectors than education is likely to be correlated with the amount of contracting out in education. The exclusion restriction is that the contracting out of non-school activities has no independent effect on student achievement in the municipality.

This requirement might not be fulfilled if the contracting out practices reflects a demand for high quality public sector services that originate from unobserved characteristics of the inhabitants. We therefore propose measures of the political composition of the local council as additional instruments. For identification it is important that the attitudes towards private schools reflect the electorate's ideological orientation and not their education orientation. Our preferred solution to this problem is to exploit the characteristics of the political party "Fremskrittspartiet" (FRP). FRP is a right wing party that strongly supports the establishment

[4] We are not able to run a direct IV estimation as equation (2) is a Tobit specification while equation (1) has a continuous dependent variable, therefore we must use a two stage approach.

of private schools. It is unlikely that this support reflects a strong education-orientation among the party's voters. In the election in 2001, 19 percent of their voters had elementary school as their highest education, 65 percent had secondary school as their highest education, while 6 percent had education at the university level. The average education level among FRP's voters is well below the national average (Aardal et al, 2003). Thus, it seems like the critical question is whether the party's support for private schools reflects a *poor* education-orientation among the FRP-voters. There is no easy answer to this question. The evidence clearly indicates the party does not represent modernisation losers in sparsely populated areas: in local elections the party achieves its best results in affluent and medium-seized municipalities, making it less obvious that a large FRP-fraction in the local council is equivalent to a poor education-orientation among the community members. There are at least two arguments why the party's positive attitudes towards private schools are not directly linked to their electorate's education orientation: First, many FRP voters hold strong anti-establishment sentiments. For this reason, they find public schools - that represent the national governments in the local community - unattractive. Second, FRP's number one priority in local policies is more money to the health sector and to care for the elderly. Privatization of public schools can be seen as a means towards this end because the schools are then taken out of the municipalities' budgets. We use the share of representatives from FRP in the local municipalities as one of our identification variables. The exclusion restriction is that this party's attitude towards private schools is not (much) correlated with their voters' educational motivations after controls are made for observable characteristics of the municipalities' inhabitants.

We also propose to use the share of representatives from the Christian Democratic Party as an identifying variable, but admittedly, we are less convinced that the exclusion restriction holds for this variable. A main reason is that religious communities may provide some kinds of social capital that may facilitate the human capital formation of young people. Finally, we exploit the geographic characteristics of Norway. A large share of the population is spread around in the fjords and up in the valleys. Most municipalities have several schools, some are located in the administrative centre, while others are established in local centres in the fjords or up in the valleys. Thus, travelling distances between the local schools and the municipal centre vary quite a lot. We conjecture that these distances effectively affect the market size for private schools, and thus the incentives for establishing such schools. The hypothesis is that fewer students go to private schools as travelling distances within the municipality increase[5]. Hastings, Kane and Staiger (2005) provide evidence that parents value proximity highly, and that travelling costs make up for much of this priority.

To check out this conjecture we use a measure on travelling distance per inhabitant. The travelling distance is calculated using a model where each municipality is divided into several smaller school districts and where the centre of the municipality and the centre of these smaller districts are defined by the most densely populated area. The travelling distances is then measured as the distance from the centre of the smaller district to the centre of the municipality. Travelling distances between local centres capture a relevant feature of the local schooling market, but are unlikely to be much correlated with the students' actual travelling time. This is important because students that spend much time travelling have less spare time for homework, indicating that actual travelling time might be a determinant of student achievement. Also, some people would argue that non-random Tiebout sorting compromise the distance measure because the most education-oriented families might have left sparsely populated areas. Note then that also many of the urban areas in Norway have relatively long travelling distances due to fjords and mountains. One illustration is that the second largest city in Norway, Bergen, which is famous for its location by fjords and among high mountains, has solved these geographical problems by allocating approximately 31 000 students to about 100 schools. We apply the assumption that conditioning both on individual student characteristics and characteristics of the municipality population, our measure of travelling distances within the municipality is not correlated with student achievement.

The socioeconomic composition of students attending private schools is somewhat different from the student body composition of public schools, also when we restrict the comparisons to municipalities with private schools. Most notably, immigrants are over-represented and also, there are relatively more students who have parents with higher educational levels (these numbers are not reported in the tables). We therefore include a selection part into the model by using Heckman's selection correction procedure.

[5] However, in Norway we have witnessed that the inhabitants in some remote areas have responded to the closing down of small public schools by establishing private schools.

The selection part of our model is formalized in the following equation:

$$(3) \qquad w_i = X_i\phi + \mu_i$$

where X_i is a vector of explanatory variables, and the dependent variable w_i can be interpreted as a student's preferences for public schooling. The latter variable is observed only if these preferences are positive in the sense that the student chooses public schooling instead of private schooling. What we observe is the binary variable:

$$(4) \qquad w_i^* = \begin{cases} 1 & if\ w_i > 0 \\ 0 & if\ w_i \leq 0 \end{cases}$$

Equation (3) is estimated using Maximum Likelihood (Probit).

As outlined so far, the analysis does not explicitly address potential problems related to Tiebout choice, i.e. that parents move home across municipality boundaries. The most education-oriented parents might have sorted themselves to municipalities with favourable peer group compositions in the public schools. (For the same reasons, the most ambitious teachers may have chosen to teach in public schools in these municipalities.) If then, these municipalities also have large shares of students in private schools, competition and sorting effects will be confounded. We deal with this problem by utilizing that the municipalities that have universities, university colleges or colleges run by the Government within their borders have the highest shares of population with higher education. This correlation will work to identify the effect of the populations' educational level on individual students' performance insofar the existence of a higher education institution in the municipality does not have any direct effects on performance.

For the present analysis the consequences of omitted variables at the municipal level might be equally important as simultaneity biases. Local decisions about school resources, school size, and teacher recruitment and so on might affect student achievement. To the extent that these variables are correlated with the private school market share, but not included among the explanatory variables, the estimated competition effect will be biased. Even though the inclusion of such choice variables generates new problems, we include a relatively large number of these variables to avoid omitted variable biases.

4. Empirical Results

4.1 Auxiliary analyses

In this chapter we present the results from the empirical analysis outlined in chapter 3. We start out by identifying the factors that determine private school location. The results come from TOBIT estimation of equation (2).

As can be seen from the table, the private school market share is positively and significantly associated with the share of children in private child care, and also with the share of representatives from the right wing party (FRP). The former association is significant at the 10 percent level, the latter at the 5 percent level. The share of representatives from the Christian Democratic Party is negatively and significantly associated with the private market share at the 10 percent level. We have also included the share of representatives from the Socialist parties. This variable is not significantly associated with the private school market share. There is evidence, significant at the 5 percent level, showing that relatively more students go to private schools in municipalities with short travelling distances from the local centres to the municipal centre.

Thus, all the suggested identifying variables are correlated with the share of students in private schools. An important question is how powerful these instruments are. A conventional way to compare coefficients is to evaluate them in terms of standard deviations. In the exercise reported below we use the standard deviations from the sub sample of municipalities that have private schools. The share of representatives from FRP varies between 0 percent and 40 percent, with a standard deviation of 6 percent. One standard deviation increase in the share of representatives from this party increases the share of students in private schools with 1.5 percent, which is about 40 percent of one standard deviation in the private school market share. One standard deviation in the contracting out of child care transforms into a 30 percent of a standard deviation increase in the share of students in private schools, while one standard deviation increase in traveling distances transforms into 45 percent of one standard deviation decrease in the share of students in private schools. Thus, the estimated coefficients for these three variables are not only statistical significant, but large enough for these variables to explain a substantial part of the variation in the private school share.

Table 4.1
First stage TOBIT regression: explaining the private school market share

The share of students in independent schools	
Market share for private kinder gardens (contracting out)	0.056 (0.031)*
The share of representatives in local council from the right wing party FRP	0.255 (0.109)**
The share of representatives in local council from the Christian Democratic Part	-0.150 (0.089)*
Travelling distance within zone per citizens	-0.004 (0.002)**
Herfindahl Index of public school competitiveness in municipality	-0.152 (0.058)***
Dummy indicating whether municipality has a University or a University College	0.016 (0.017)
Herfindahl Index of educational level in the municipality	-0.356 (0.334)
The share of employees between 20 and 54 in the municipality	-0.567 (0.155)***
The share of immigrants in the municipality	-0.041 (0.311)
Population size (in 1000)	0.000 (0.000)
The school cost per student in the municipality	0.001 (0.001)
Municipal revenue	-0.000 (0.001)
The inverse number of students	2.006 (7.301)
The share of socialists in the local council	-0.040 (0.051)
Constant	0.323 (0.145)**
Observations	424

Note: All estimation results are reported with heteroscedasticity robust standard errors in parenthesis. *, ** and *** indicating significance at 10 %, 5 % and 1 % respectively.

It is useful to know whether private schools are established as a response to high demand for educational quality and/or as a response to poor public school quality. The equation reported in Table 4.1 includes a dummy reflecting whether the school district has a university or one or more University Colleges. This variable, which might be an indicator of the inhabitants' education-orientation, is not significantly associated with the dependent variable. In an alternative specification (not reported) we have included the share of the population with higher education. This variable is a significant predictor of private school location, indicating that there is some imperfect Tiebout sorting across municipalities.

Many authors hypothesize that the demand for private schools depends on the quality of the public schools. Sandstrøm and Bergstrøm (2005) have access to the average grades in public schools in the municipalities at the time of the introduction of the new private school law in

Sweden. Thereby they are able to provide indications that the establishment of independent schools in Sweden has been a response to poor public school performance. This result indicates that they run the risk of underestimating the positive effects of competition if endogeneity is not taken into account. We have no adequate measures for public school performance at the time of the establishment of private schools. Instead, we include current expenditures per student (which due to its low variability over time correlates strongly with historical inputs). This is a relevant quality measure if parents judge schools by inputs. Expenditures per student in public schools are not significantly related to the location of private schools. We return to the relationship between public school quality and the establishment of private schools below.

Many of the other independent variables in Table 4.1 are motivated by prior empirical studies. Glomm, Harris and Lo (2005) provide evidence from Michigan that charter schools locate where populations are diverse in terms of race and adult education levels. We include a Herfindahl index for years of schooling and the share of students with immigrant background. None of these variables appear to have a significant impact on the establishment of private schools in Norway. Figlio and Stone (2001) hypothesize that the demand for private schools depends on the choice options within the public sector, and add the number of public school options to the list of explanatory variables. We use a Herfindahl Index as our measure of current public school concentration, and find that there is significant evidence that more students go to private schools in municipalities with low public school concentration. This result, which is contrary to the findings reported by Figlio and Stone (2001), indicates that the mechanism underlying the establishment of private schools in Norway does not seem to be the lack of competition between public schools. Finally, note that the private school market share seems to be smaller in municipalities where a large fraction of the inhabitants are between 20 and 54 years. A similar finding is reported by Barrow (forthcoming), who finds that religious schools in Illinois are more likely to locate in areas with a higher share of the population over 55 years of age. According to Barrow this correlation might reflect past location decisions.

Table 4.2
The selection equation (Probit-estimation)

Dummy for girls	-0.176 (0.054)***
First generation immigrant with no Norwegian background	0.073 (0.183)
Born in Norway by two foreign parents	-0.422 (0.143)***
Born abroad with one Norwegian parent	-0.471 (0.294)
Born in Norway with one foreign parent	-0.263 (0.070)***
Born abroad by two Norwegian parents	-0.536 (0.162)***
Father's education is lower tertiary or higher (College or university)	-0.452 (0.114)***
Father's education is upper secondary	-0.215 (0.103)**
Mother's education is lower tertiary or higher (College or university)	-0.975 (0.132)***
Mother's education is upper secondary	-0.584 (0.114)***
Father's income (in 10 000 NOK)	0.002 (0.001)***
Mother's income (in 10 000 NOK)	0.006 (0.001)***
Living in nuclear family	-0.199 (0.066)***
Birth order	0.178 (0.034)***
Number of siblings	-0.229 (0.042)***
Mother works in private sector	-0.108 (0.051)**
Father works in private sector	-0.060 (0.038)
Number of teacher hours per student at the school	0.005 (0.005)
Share of students receiving extra Norwegian tuition at the school	-1.615 (1.556)
Share of students receiving special training (disadvantaged students)	8.964 (4.225)**
School size (in 100)	0.985 (0.274)***
The share of teacher with certified education	13.760 (2.456)***
Dummy indicating whether municipality has a University or a University College	-0.661 (0.378)*
Herfindahl Index of educational level in municipality	23.671 (10.568)**
The share of employees between 20 and 54 municipality	0.816 (4.296)
The share of immigrants in municipality	11.928 (6.130)*
Population size (in 1000)	-0.007 (0.002)***
The school cost per student in municipality	-0.022 (0.032)
Municipal Revenue	0.096 (0.040)**
The share of representatives in local council from the right wing party FRP	-13.244 (3.833)***
The share of representatives in local council from the Christian Democratic Part	2.813 (3.583)
The share of socialists in local council	-1.540 (1.837)
Constant	-22.087 (6.983)***
Observations	48401

Note: All estimation results are reported with heteroscedasticity robust standard errors, adjusted for school level clustering, in parenthesis. * and ** indicating significance at 5 % and 1 % respectively.

The characteristics of the students who choose to stay in the public schools, and who choose to attend private schools, are highlighted in table 4.2, where we report the PROBIT estimation results of equation (3)[6]. These results show that girls are more likely to choose private schools than are boys, and that students with highly educated parents are more likely to choose private schools than are students with less educated parents. Later on we investigate whether the school competition effects differ across these two subgroups of students. Note that several other individual student and family characteristics – such as immigrant status, family income, family size and family structure, are significant predictors of the choice of private versus public schools as well.

We have also included variables at the municipal level in this equation. It should be noted that the probit results differ from the Tobit results reported in Table 4.1 in important respects. Examples are that the probit specification shows that students in municipalities with Universities or University Colleges are more likely to opt out for private schools, and that the diversity of adult education levels seems to trigger private school enrolment (neither is the case in the Tobit specification). Diverging results basically reflect that the probit estimation makes use individual data that represent a different weighting of the observations

A potentially important result is that municipalities with a large share of public school teachers without certification experience a larger outflow of students to private schools. Municipalities with uncertified teachers are in short supply of teachers, and are thus less able to choose high quality teachers than are municipalities that experience excess teacher supply. The share of uncertified teachers might be an indicator of the average teacher quality in public schools, and thus the strongest evidence provided in the present paper that students are more likely to opt out of poor public schools. For the selection equation to be useful in a Heckman-correction procedure, it should contain at least one variable measured at the municipality level, and that is not included in the education function or the Tobit equation (1). Unfortunately, we have been unable to include any variables that fulfill these requirements into the selection equation. Identification thus basically hinges on the functional form. This should be kept in mind when we now turn to the estimations of the education production function.

---

[6] The sample that is used in this estimation is based all 10th grade students that has valid test results in at least one of the three test subjects, the number of observations in table 5.2 is thus larger than in any of the preceding estimations based on one subject at the time. Due to missing observation on individual background variables the number of observations is also somewhat lower than the total number of public school 10th graders in the data.

4.2 The education production function analyses

The results from estimating equation (1) are presented in tables 4.3-4.5[7]. For each of the three subjects we report results from both ordinary least squares and instrumental variables estimation. All estimates are presented using robust standard errors allowing for clustering at the school level. The ordinary least squares estimations reported in columns (1) provide a useful point of departure. Reading across the three tables, the competition effects from private schools are positive and insignificant for mathematics, while the comparable estimates for reading and writing and English are negative and insignificant.

Table 4.3
Competition effects in mathematics

Dependent variable:	Test scores in Mathematics					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	0.654	8.586	1.137	16.014	0.721	5.679
	(3.530)	(3.340)**	(4.883)	(4.569)***	(3.949)	(3.709)
Inverse Mills Ratio	-0.294	-0.649	1.023	0.260	0.737	0.507
	(0.881)	(0.883)	(1.387)	(1.362)	(1.239)	(1.219)
Herfindahl Index of public school	-0.279	1.317	-1.352	1.907	0.328	1.308
competitiveness in municipality	(0.708)	(0.901)	(0.985)	(1.288)	(0.786)	(0.992)
Constant	30.945	28.446	30.844	26.814	36.368	34.493
	(3.976)***	(4.095)***	(5.578)***	(5.596)***	(4.246)***	(4.428)***
Observations	40237	40234	16868	16867	23369	23367
R-squared	0.17	0.17	0.06	0.06	0.04	0.04

Note: All estimation results are reported with heteroscedasticity robust standard errors, adjusted for school level clustering, in parenthesis. * and ** indicating significance at 5 % and 1 % respectively.

The ordinary least square estimations reported in the tables include the Inverse Mills Ratio (originating from equation (3)). This variable has a negative and insignificant coefficient for all three subjects. The negative signs of the Inverse Mills Ratios indicate that there are some unobserved variables that on one hand affect student achievement positively, and on the other hand affect the probability of choosing public schools negatively[8]. In other words, we find some insignificant indications of self-selection into public schools by students with relatively poorer education motivation. The correction for self-selection does not affect the estimated competition effects much, probably reflecting that the Norwegian private school market is small (the specifications without the Inverse Mills Ratio are not reported).

[7] All the presented estimates emerge from a model where we include an extensive list of control variables at the individual, school and municipality level. In table A4 in the appendix we present the estimation results in mathematics where all control variables are included in the model.
[8] The coefficient on the Inverse Mills Ratio will generally take on the same sign as the correlation (rho) between the error term in the selection equation and the error term in the achievement equation.

In the second columns of tables 4.3-4.5 we present results from IV estimations (we use the predicted private school market share from the TOBIT equation reported in table 4.1 for the observed private school market shares). The private school market share now exercises a positive and significant influence on public school performance in mathematics. A four percentage point increase in the private school market share (which is a little more than one standard deviation in the market share for the municipalities that have private schools) transforms into 3.4 points on the math test, which is 3.4 percent of a standard deviation in the test results for mathematics. In their analysis of Swedish lower secondary schools, Sandstrøm and Bergstrøm (2005) estimate the competition effect to be more than twice this size. However, they use an outcome measure that is a composite of the students' grades in 16 subjects. It is thus not obvious that these numbers are comparable.

The estimated IV-coefficient in table 4.3 is considerably larger than the coefficient originating from the OLS estimation, indicating that the competition effect is seriously underestimated unless simultaneity is taken into account. The direction of the difference between the OLS- and IV-coefficients indicates that private schools are established in municipalities where public schools show poor performance in mathematics.

Table 4.4
*Competition effects in Reading and writing**

Dependent variable:	Test scores in Reading and Writing					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-2.007	-0.082	-0.314	6.810	-3.022	-2.508
	(4.046)	(2.781)	(4.039)	(3.237)**	(4.954)	(3.557)
Inverse Mills Ratio	-0.636	-0.635	0.726	0.387	0.451	0.566
	(0.793)	(0.799)	(0.908)	(0.903)	(0.977)	(0.971)
Herfindahl Index of public school competitiveness	0.567	0.577	0.901	2.260	0.199	-0.204
in municipality	(0.626)	(0.826)	(0.758)	(0.967)**	(0.769)	(1.020)
Constant	31.359	31.215	36.912	34.946	32.694	33.229
	(3.254)***	(3.325)***	(4.057)***	(4.077)***	(3.921)***	(4.068)***
Observations	41535	41532	17280	17279	24255	24253
R-squared	0.17	0.17	0.08	0.08	0.08	0.08

* see note Table 4.3.

The IV estimation results reported in the second columns of Tables 4.4 and 4.5 indicate that there are no uniform positive competition effects across subjects. In English the estimated coefficient is negative and insignificant. For reading and writing the estimated IV-coefficient is close to zero and also insignificant. Note also that both for English and for reading and writing, the coefficients estimated by the use of IV are larger than the coefficient estimated by OLS, implying that the biases in the OLS coefficients are consistent across subjects.

The specifications reported in the two first columns in tables 4.3-4.5 include all students, and conceal potentially important differences across student subgroups. In the introduction, we have conjectured that teachers respond to more competition pressures by concentrating more of their efforts on the student subgroups that are most likely to leave public schools for private schools. Hastings, Kane and Staiger (2005) provide empirical evidence that parental preferences for school quality vary with background characteristics in ways that are consistent with this conjecture. Table 4.2 shows that the probability of attending a private school in Norway depends both on gender and parental education. Consequently, we split the sample along these lines.

Table 4.5
*Competition effects in English**

Dependent variable:	Test scores in English					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-9.098	-3.398	-11.592	-2.297	-8.096	-2.326
	(4.981)*	(3.418)	(5.724)**	(4.476)	(5.422)	(3.595)
Inverse Mills Ratio	-1.703	-1.563	-1.014	-0.975	-0.470	-0.344
	(1.384)	(1.395)	(1.615)	(1.640)	(1.349)	(1.350)
Herfindahl Index of public school	0.301	-0.211	0.705	0.446	-0.231	-0.551
competitiveness in municipality	(0.826)	(0.972)	(1.182)	(1.407)	(0.828)	(1.020)
	(4.109)***	(4.348)***	(5.319)***	(5.519)***	(4.273)***	(4.552)***
Observations	41023	41020	17252	17251	23771	23769
R-squared	0.14	0.14	0.05	0.05	0.06	0.06

* see note Table 4.3.

First, the results reported in the right hand parts of Table 4.3 (mathematics) and Table 4.4 (reading and writing) show that public school students with highly educated parents gain from competition with private schools. The estimated effects are highly significant when using instrumental variables estimation. The estimated effect in mathematics is more than twice the size of the effect in reading and writing. Students with highly educated parents seem to experience no positive effect from competition in English. There is no significant evidence that students with less educated parents are affected by competition – anywhere.[9] Both for mathematics and reading and writing the hypothesis that the competition effects are equal for students with highly educated parents and students with less educated parents is rejected at conventional levels of significance. Thus, these findings are consistent with the hypothesis

[9] We have also investigated whether the positive competition effects are larger in densely populated municipalities: they are, and more evident for mathematics than for reading. These results are not reported in the tables.

that public schools respond to increased competition by allocating their efforts to the students that are most likely to leave the public schools.

Next, the sample is separated by gender. The results are reported in Tables 4.6 and 4.7 for boys and girls respectively. The positive competition effects in mathematics seem to originate from boys' responses. Boys seem more than twice as responsive as girls, and the competition effect is significant for boys only. Recall from Table 4.2 that girls, all else equal, are more likely to choose private schools. The hypothesis that teachers respond to increased competition by concentrating on the student subgroups that are most likely to leave public schools are thus not supported by these findings[10]. In reading and writing there are some weak indications (insignificant at conventional levels) that boys might be more responsive than girls also in this subject. In English we find no significant effects.

Table 4.6
*Competition effects for boys**

Dependent variable:	Test scores in Mathematics		Test scores in English		Test scores in Reading and writing	
Sample:	Boys					
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-1.800	11.748	-10.364	-1.285	-2.216	1.838
	(4.576)	(3.850)***	(6.095)*	(4.029)	(5.156)	(3.726)
Inverse Mills Ratio	0.931	0.489	-0.752	-0.685	-0.457	-0.531
	(1.170)	(1.153)	(1.621)	(1.635)	(0.975)	(0.985)
Herfindahl Index of public school	-0.976	1.249	-1.047	-1.160	-0.229	0.140
competitiveness in municipality	(0.857)	(1.080)	(0.994)	(1.198)	(0.846)	(1.132)
Constant	26.256	22.419	33.314	32.637	28.883	28.095
	(4.661)***	(4.884)***	(4.657)***	(4.932)***	(4.217)***	(4.439)***
Observations	20587	20585	20768	20766	21252	21250
R-squared	0.17	0.17	0.12	0.12	0.15	0.15

* see note Table 4.3.

All the specifications reported above include a large number of control variables at the individual level, the school level and the municipal level. In the present context, the Herfindahl-Hirschman index that is used to measure competition between public schools within municipalities warrant some comments. This index is calculated as the sum of each public school's squared market share (share of total number of students in public schools in the municipality). An index of 1 means that all students attend the same public school and that there are no competition between public schools in the municipality, while an index of 0.5 means that the students in public schools are equally distributed into two competing public schools. The closer the index is to zero the stronger is the public school dispersion in the

[10] However, these findings are consistent with a small number of empirical analyses showing that schools matter more for boys than for girls.

municipality. This is potentially an important control variable because the private school market share is significantly larger in municipalities associated with low public school concentration (see table 4.1). Hoxby (2000) and Hanushek and Rivkin (2003) find in their studies that more competition among public schools (less market concentration) is associated with higher educational outcomes and higher teacher quality. Omitting the measure of public school competition from the estimated equations might thus generate an upward bias in the estimated competition effects from private schools.

We are not able to report any clear evidence of how the potential competition between public schools affects public schools performance. The estimated effects are close to zero and insignificant for all three subjects except for the sub sample of students with highly educated parents, where we find a significant and positive effect in reading and writing. In this case the estimated coefficient of competition between public schools is about one third of the private competition coefficient. We would expect that the competition between public schools is weaker than between public and private schools due to institutional differences (neighborhood public schools). However, the value of this exercise is limited by the fact that public school concentration is likely to be endogenous to public school performance. We do not deal with this potential simultaneity problem in this paper.

Table 4.7
*Competition effects for girls**

Dependent variable:	Test scores in Mathematics		Test scores in English		Test scores in Reading and writing	
Sample:	Girls					
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	3.104	4.974	-7.657	-5.725	-1.776	-2.163
	(4.249)	(3.916)	(5.122)	(3.842)	(4.178)	(3.098)
Inverse Mills Ratio	-1.400	-1.617	-2.329	-2.101	-0.454	-0.362
	(1.096)	(1.101)	(1.493)	(1.495)	(0.825)	(0.820)
Herfindahl Index of public school competitiveness in	0.520	1.402	1.694	0.751	1.426	1.055
municipality	(0.852)	(1.100)	(0.967)*	(1.154)	(0.706)**	(0.926)
Constant	35.932	34.717	46.352	47.596	36.790	37.339
	(4.732)***	(4.854)***	(5.113)***	(5.312)***	(3.721)***	(3.786)***
Observations	19650	19649	20255	20254	20283	20282
R-squared	0.17	0.17	0.11	0.11	0.15	0.15

* see note Table 4.3.

5. Robustness Checks

The positive effects of competition from private schools, as reported above, depend on the chosen specifications. Two of our choices are of particular importance. The first important choice is to use the results from the national tests as our measure of student performance. As stated above, we also have access to the teachers' grades and the students' exam results. When the same models as presented in chapter 4 are estimated using the teachers' grades as the measure on student achievement, our previous findings are confirmed to some extent only. The analyses indicate small and insignificant effects from private competition in all subjects when all students are included in the estimated equations. Estimating the equation for the same student subgroups as above, we find a positive and significant effect from private competition for boys in mathematics, but no significant effects for any of the other sub groups. The estimated competition effect for boys' grades in math is only half the size of the effect found using national test results for the same boys. The difference in competition effects between the highlighted subgroups of students are not affected by the choice of outcome measures.

We argue that the results based on the teachers' grades are less credible. The evidence supporting this claim comes from equations where the teachers' grades in math are regressed against the students' exam results in math (using the sub sample of students that have taken the math exam) while including all the independent variables from the education production function equation. The estimation is reported in table 5.1 and provides weak evidence that the teachers' grading practices are more lenient in municipalities with a large share of students in private schools. On the other hand, the grading practices are significantly more restrictive in municipalities characterized by a large share of highly educated inhabitants and where the share of certified teachers is large. Thus, this exercise provides clear evidence that the teachers' grading practices reflect environmental characteristics in systematic ways, making the grades less credible measures of student performance.

Our second important choice is to exclude the share of the municipal population with higher education from the estimated equations. To see the implications, we have estimated our model using the endogenous share of municipal population with higher education both in the first stage regressions and in the education production functions. These regressions; which are reported in table 5.2 below for mathematics, show that the estimated competition effects are sensitive to the inclusion of this variable[11]. We see that the share of highly educated people constitutes a positive and highly significant effect in mathematics and reading and writing. Due to the strong correlation with the private market share, the estimated private competition

[11] Estimation results for English and reading and writing are presented in the Appendix in tables A5 and A6 respectively.

effects weaken substantially. The share of highly educated people does not have any noticeable effect on student achievement in English, and consequently we only observe small changes in the private competition effects for this subject. The inclusion of the share of the municipal population with higher education does not affect the difference in responses from students with highly educated parents and students with less educated parents. The latter findings are not reported in the tables.

Table 5.1
*Testing for Grade Inflation from competition in mathematics**

Dependent variable:	Assessment grades in Mathematics	
	OLS	OLS
External examination score in Mathematics	0.840 (0.005)***	0.809 (0.005)***
The share of students in private schools		9.032 (4.139)**
Inverse Mills Ratio		1.271 (1.315)
Herfindahl Index of public school competitiveness in municipality		-1.628 (0.861)*
Number of teacher hours per student at the school		-0.004 (0.006)
Share of students receiving extra Norwegian tuition at the school		-0.494 (0.237)**
Share of students receiving special training (disadvantaged students)		1.141 (2.759)
School size (in 100)		-0.106 (0.145)
The share of teacher with certified education		-6.173 (2.772)**
The share of population with higher education		-4.185 (2.364)*
The share of employees between 20 and 54 municipality		-0.866 (3.930)
The share of immigrants in municipality		2.972 (5.181)
Population size (in 1000)		-0.001 (0.002)
The school cost per student in municipality		0.051 (0.019)***
Municipal Revenue		0.015 (0.019)
Constant	8.024 (0.279)***	11.322 (4.259)***
Observations	19751	18444
R-squared	0.71	0.72

* see note Table 4.3

An important question is how the strong relationship between public school students' performance and the share of highly educated people in the municipality should be interpreted. Above we have argued that the strong correlation between public school student performance and the municipal educational level might reflect that people with a strong engagement for schooling have sorted themselves into certain municipalities, and moreover that a high level of educated people is attractive for people that are generally concerned about schooling. In the analyses reported above, steps are taken to deal with this problem.

There are other interpretations to consider. It is well documented that parental education is one of the most important determinants of student achievement, but we claim that it is unlikely that the educational level in the municipality has a strong direct effect on student achievement after controlling for parental education at the individual level. If there is a peer-group effect present, it should be more precisely captured at the school, or class level.

Table 5.2
*Robustness checks using the endogenous share of population with higher education**

Dependent variable:	Test scores in Mathematics					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-3.680	-1.251	-5.446	2.787	-3.720	-5.074
	(3.576)	(3.858)	(5.062)	(5.518)	(3.830)	(4.142)
Inverse Mills Ratio	-0.508	-0.464	0.601	0.408	0.674	0.912
	(0.923)	(0.935)	(1.405)	(1.418)	(1.318)	(1.342)
Herfindahl Index of public school competitiveness in municipality	0.210	0.026	-0.456	0.125	0.600	-0.227
	(0.696)	(0.940)	(0.950)	(1.323)	(0.778)	(1.040)
The share of population with higher education	10.690	10.740	14.316	12.867	12.292	13.615
	(2.055)***	(2.403)***	(2.594)***	(3.158)***	(2.409)***	(2.719)***
Herfindahl Index of educational level in municipality	20.780	20.733	28.624	30.161	21.885	20.600
	(6.715)***	(6.922)***	(9.116)***	(9.433)***	(7.821)***	(7.969)***
The share of employees between 20 and 54 municipality	0.556	0.038	1.923	3.771	-0.496	-3.229
	(2.886)	(3.606)	(4.068)	(5.041)	(3.139)	(3.911)
The share of immigrants in municipality	6.915	6.772	11.230	10.491	3.669	3.949
	(4.605)	(4.596)	(5.983)*	(5.949)*	(4.929)	(4.934)
Population size (in 1000)	0.001	0.001	0.001	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
The school cost per student in municipality	0.037	0.035	0.036	0.029	0.043	0.042
	(0.015)**	(0.015)**	(0.022)*	(0.022)	(0.017)**	(0.017)**
Municipal Revenue	-0.021	-0.020	-0.033	-0.027	-0.025	-0.027
	(0.016)	(0.016)	(0.022)	(0.022)	(0.017)	(0.017)
Constant	32.842	32.952	34.501	33.223	37.371	38.746
	(3.792)***	(4.017)***	(5.163)***	(5.560)***	(4.122)***	(4.356)***
Observations	40237	40234	16868	16867	23369	23367
R-squared	0.17	0.17	0.07	0.07	0.05	0.05

* see note Table 4.3.

Unfortunately, our data does not allow for linking students to specific classrooms[12], but we have tried out a specification of our model including the average educational level among parents at the school level[13]. Using this specification of the model does not change our main results and moreover, after controlling for individual background characteristics we find no effect on student achievement from the average parental educational level in schools. Other possible explanations for the large municipality effects we find in tables 5.2 and 5.3 might be that municipalities with highly educated population more easily attract competent teachers, or that the school owners might actively increase their effort as a response to educational oriented inhabitants who (beyond the demands from parents) communicate high expectation and demand more from the local schools in terms of quality and results. These interpretations raise potentially important questions both about teacher allocations across municipalities and voice mechanisms. Unfortunately, no data that can be used to shed light on these issues has been available for the present study.

6. Concluding Remarks

A heated policy discussion about the pros and cons of school competition takes place in many countries. The present study contributes to this discussion by investigating whether the performance of Norwegian public schools is affected by competition from "Norwegian type" private schools. A robust finding is that public school students with less educated parents do not realize any positive effects from this kind of school competition. For students with highly educated parents it is harder to pin down the exact effects. Specifications that include the share of population with higher education in the municipality show a positive, but small and insignificant, effect of competition for this student subgroup. Specifications that exclude the share of population with higher education show a positive, quite large, and highly significant effect of competition for this subgroup. We suspect that neither of these specifications provides a correct estimate of the competition effect, but instead, these estimations define the range for the competition effects. Importantly, and that is our most basic finding, the difference in competition effects across the two subgroups of students is very robust to alternative specifications.

[12] It is off course possible for the schools that only have on class on each grade, but this represents only a fraction of the student population.
[13] This variable is measured as the average parental educational level for each individual's peers at the school.

We argue that the heterogeneous competition effects across student subgroups should be interpreted together with the evidence that students with highly educated parents are much more likely to choose private schools than are students with less educated parents. Our findings might thus indicate that public school actors respond to competition by reallocating resources or attention from less attractive students who are likely to stay in public schools to attractive students who are more likely to leave public schools. It is realized that the "black box" approach to public schools makes this a speculative interpretation. One of the interesting issues left for further research is whether ability tracking is applied to a larger extent by public schools that are located in the most competitive environments.

References

Aardal, B., Valen, H., Karlsen, R., Kleven. Ø. and Normann, T.M. (2003): Valgundersøkelsen 2001, Report 2003/14, Statistics Norway

Barrow, L. (2007): Private School Location and Neighbourhood Characteristics, Economics of Education Review, forthcoming

Bayer, P. and McMillan R. (2005): Choice and Competition in Local Education Markets, Working paper No. 11802, National Bureau of Economic Research.

Bettinger, E. (2005): The Effect of Charter Schools on Charter Students and Public Schools, Economics of Education Review, 24, pp. 133-147

Card, D., Dooley, M. and Payne, A.A. (2006): School Competition with Publicly-Funded Catholic Schools: Evidence on Student Mobility and Achievement, Unpublished manuscript.

Dee, T. S., (1998): Competition and the Quality of Public Schools, Economics of Education Review, 17, pp. 419-427

Figlio, D.N. and Stone, J.A. (2001): Can Public Policy Affect Private School Cream Skimming?, Journal of Urban Economics, 49, pp. 240-266.

Fiske, E.B. and Ladd H. (2001): Does Competition Generate Better Schools? Evidence from New Zealand, Working Paper Series SAN01-16, Terry Sandford Institute of Public Policy.

Gibbons, S, S. Machin, S and Silva, O. (2005): Competition, Choice and Pupil Achievement, Working paper, Centre for the Economics of Schooling, London School of Economics.

Glomm, G., Harris, D. and Lo, Te-Fen. (2005): Charter School Location, Economics of Education Review, 24, pp. 451-457.

Hanushek E. A. (2002): Publicly Provided Education, In: Auerbach, A.J., Feldstein, M. (Eds.), Handbook in Public Economics. Amsterdam: North-Holland, pp. 2045-2141.

Hanushek, E. A. and Rivkin, S.G. (2003): Does Public School Competition Affect Teacher Quality?, In Hoxby, C.M. (ed.), The Economics of School Choice. Chicago: University of Chicago Press, pp. 23-47

Hastings, J.S., Kane, T.J. and Staiger, D.O. (2005): Parental Preferences and School Competition: Evidence from a Public School Program, Working paper no. 11805, National Bureau of Economic Research.

Hoxby, C.M. (1994): Do Private Schools Provide Competition for Public Schools?, Working Paper no. 4978, National Bureau of Economic Research.

Hoxby, C.M. (2000): Does Competition Among Public Schools Benefit Students and Taxpayers? The American Economic Review, 90, pp. 1209-1238.

Hsieh, C.T. and Urquiola, M. (2006): The Effects of generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program, Journal of Public Economics, 90, pp. 1477-1503

Levačić, R. (2004): Competition and the Performance of English Secondary Schools: Further Evidence, Education Economics, 12, pp. 177-193.

Newmark, C. M. (1995): Another Look at Whether Private Schools Influence Public School Quality, Public Choice, 82, pp. 365-373.

Sandstrøm, F.M., and Bergstrøm, F. (2005): School Vouchers in Practice: Competition Will Not Hurt You, Journal of Public Economics, 89, pp. 351-380.

Wikström, C., and Wikström, M. (2005): Grade Inflation and School Competition: an Empirical Analysis Based on the Swedish Upper Secondary Schools, Economics of Education Review, 24, pp. 251-267

Appendix

Table A1
Mathematics

Share of students attending private schools in the municipality	Number of students in the municipalities	Mean test score
No students	21 534	49.44
Between zero and 2 %	3 410	49.48
Between 2 % and 6 %	16 329	51.24
Between 6 % and 10 %	1 996	49.92
Above 10 %	525	49.70

Table A2
Reading and writing

Share of students attending private schools in the municipality	Number of students in the municipalities	Mean test score
No students	22 435	49.59
Between zero and 2 %	3 427	49.76
Between 2 % and 6 %	17 133	51.20
Between 6 % and 10 %	1 930	50.20
Above 10 %	442	49.74

Table A3
English

Share of students attending private schools in the municipality	Number of students in the municipalities	Mean test score
No students	21 753	49.69
Between zero and 2 %	3 410	49.33
Between 2 % and 6 %	17 077	50.86
Between 6 % and 10 %	2 007	49.96
Above 10 %	514	48.63

Table A4

*Competition effects in mathematics: full list of control variables**

Dependent variable:	Test scores in Mathematics					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	0.654	8.586	1.137	16.014	0.721	5.679
	(3.530)	(3.340)**	(4.883)	(4.569)***	(3.949)	(3.709)
Inverse Mills Ratio	-0.294	-0.649	1.023	0.260	0.737	0.507
	(0.881)	(0.883)	(1.387)	(1.362)	(1.239)	(1.219)
Herfindahl Index of public school competitiveness in municipality	-0.279	1.317	-1.352	1.907	0.328	1.308
	(0.708)	(0.901)	(0.985)	(1.288)	(0.786)	(0.992)
Dummy for girls	-0.173	-0.176	-0.367	-0.377	-0.056	-0.056
	(0.109)	(0.110)	(0.162)**	(0.162)**	(0.131)	(0.132)
First generation immigrant with no Norwegian background	-3.170	-3.194	-4.924	-4.963	-2.871	-2.884
	(0.404)***	(0.403)***	(0.769)***	(0.768)***	(0.445)***	(0.445)***
Born in Norway by two foreign parents	-1.011	-0.993	-1.869	-1.841	-1.210	-1.189
	(0.394)**	(0.396)**	(0.671)***	(0.671)***	(0.456)***	(0.456)***
Born abroad with one Norwegian parent	-1.497	-1.562	-1.372	-1.419	-2.219	-2.310
	(0.718)**	(0.715)**	(0.983)	(0.979)	(1.174)*	(1.168)**
Born in Norway with one foreign parent	-0.884	-0.907	-1.223	-1.261	-0.418	-0.432
	(0.213)***	(0.212)***	(0.300)***	(0.299)***	(0.311)	(0.310)
Born abroad by two Norwegian parents	-3.882	-3.872	-4.266	-4.259	-2.684	-2.675
	(0.491)***	(0.491)***	(0.668)***	(0.667)***	(0.662)***	(0.662)***
Father's education is lower tertiary or higher (College or university)	5.822	5.796				
	(0.218)***	(0.217)***				
Father's education is upper secondary	2.081	2.071				
	(0.151)***	(0.151)***				
Mother's education is lower tertiary or higher (College or university)	5.732	5.723				
	(0.205)***	(0.206)***				
Mother's education is upper secondary	2.379	2.378				
	(0.165)***	(0.165)***				
Father's income (in 10 000 NOK)	0.012	0.012	0.012	0.012	0.033	0.033
	(0.005)***	(0.004)***	(0.006)**	(0.006)**	(0.005)***	(0.005)***
Mother's income (in 10 000 NOK)	0.037	0.037	0.056	0.056	0.057	0.057
	(0.004)***	(0.004)***	(0.005)***	(0.005)***	(0.006)***	(0.006)***
Living in nuclear family	2.885	2.882	3.356	3.342	2.781	2.782
	(0.103)***	(0.103)***	(0.186)***	(0.185)***	(0.130)***	(0.130)***
Birth order	0.532	0.527	0.663	0.652	0.483	0.481
	(0.076)***	(0.076)***	(0.116)***	(0.116)***	(0.099)***	(0.099)***
Number of siblings	-0.081	-0.074	0.097	0.109	-0.101	-0.096
	(0.074)	(0.074)	(0.117)	(0.117)	(0.094)	(0.094)
Mother works in private sector	0.570	0.558	0.257	0.219	0.419	0.415
	(0.100)***	(0.100)***	(0.166)	(0.166)	(0.129)***	(0.129)***
Father works in private sector	0.229	0.224	-0.418	-0.436	0.055	0.058
	(0.125)*	(0.124)*	(0.171)**	(0.170)**	(0.147)	(0.147)
Number of teacher hours per student at the school	0.004	0.005	0.011	0.012	-0.002	-0.001
	(0.007)	(0.007)	(0.011)	(0.010)	(0.008)	(0.008)
Share of students receiving extra Norwegian tuition at the school	-0.257	-0.121	-1.231	-0.888	-0.012	0.062
	(1.065)	(1.083)	(1.806)	(1.809)	(0.933)	(0.945)
Share of students receiving special training (disadvantaged students)	-4.157	-4.283	-2.459	-2.863	-5.685	-5.743
	(2.782)	(2.763)	(3.577)	(3.538)	(3.047)*	(3.025)*
School size (in 100)	-0.019	-0.052	0.137	0.067	-0.032	-0.053
	(0.089)	(0.089)	(0.123)	(0.121)	(0.097)	(0.098)
The share of teacher with certified education	2.973	2.361	7.975	6.757	2.126	1.711
	(2.571)	(2.572)	(3.849)**	(3.735)*	(2.679)	(2.697)
Dummy indicating whether municipality has a University or a University College	-0.241	-0.290	-0.343	-0.360	-0.097	-0.147
	(0.245)	(0.243)	(0.318)	(0.307)	(0.269)	(0.270)
Herfindahl Index of educational level in municipality	2.463	9.442	0.164	14.883	3.460	7.769
	(6.044)	(6.555)	(9.056)	(9.449)	(6.782)	(7.407)
The share of employees between 20 and 54 municipality	4.274	8.364	8.030	15.148	3.094	5.964
	(2.819)	(3.159)***	(3.989)**	(4.296)***	(3.108)	(3.534)*
The share of immigrants in municipality	7.044	5.806	10.335	8.314	4.451	3.651
	(4.643)	(4.606)	(6.106)*	(5.984)	(4.993)	(4.981)
Population size (in 1000)	0.002	0.002	0.002	0.002	0.002	0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
The school cost per student in municipality	0.023	0.024	0.012	0.011	0.031	0.032
	(0.016)	(0.016)	(0.024)	(0.022)	(0.017)*	(0.017)*
Municipal Revenue	0.013	0.010	0.026	0.017	0.004	0.004
	(0.018)	(0.017)	(0.024)	(0.022)	(0.018)	(0.018)
Constant	30.945	28.446	30.844	26.814	36.368	34.493
	(3.976)***	(4.095)***	(5.578)***	(5.596)***	(4.246)***	(4.428)***
Observations	40237	40234	16868	16867	23369	23367
R-squared	0.17	0.17	0.06	0.06	0.04	0.04

* see note Table 4.3.

Table A5

*Robustness checks using the endogenous share of population with higher education**

Dependent variable:	Test scores in Reading and Writing					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-5.318	-8.444	-5.630	-4.939	-6.850	-12.304
	(3.968)	(3.234)***	(4.001)	(3.735)	(4.795)	(3.979)***
Inverse Mills Ratio	-0.734	-0.363	0.659	0.834	0.255	0.831
	(0.826)	(0.821)	(0.947)	(0.952)	(1.144)	(1.132)
Herfindahl Index of public school competitiveness in municipality	0.975	-0.430	1.560	0.730	0.558	-1.447
	(0.613)	(0.853)	(0.714)**	(0.954)	(0.755)	(1.065)
The share of population with higher education	7.122	9.286	11.137	12.149	8.758	11.977
	(1.709)***	(1.988)***	(1.960)***	(2.338)***	(2.279)***	(2.543)***
Herfindahl Index of educational level in municipality	2.595	-0.135	12.442	11.221	6.901	2.778
	(6.442)	(6.446)	(7.755)	(7.795)	(7.697)	(7.704)
The share of employees between 20 and 54 municipality	5.968	1.205	3.509	0.778	5.979	-0.922
	(2.733)**	(3.325)	(3.301)	(3.809)	(3.145)*	(4.042)
The share of immigrants in municipality	6.363	6.815	7.439	7.552	7.094	7.809
	(3.576)*	(3.594)*	(3.945)*	(3.965)*	(4.558)	(4.586)*
Population size (in 1000)	0.002	0.002	0.002	0.002	0.001	0.001
	(0.001)*	(0.001)	(0.001)*	(0.001)	(0.002)	(0.002)
The school cost per student in municipality	0.051	0.051	0.037	0.035	0.062	0.062
	(0.015)***	(0.015)***	(0.018)**	(0.018)*	(0.018)***	(0.018)***
Municipal Revenue	-0.025	-0.028	-0.046	-0.045	-0.020	-0.024
	(0.014)*	(0.014)**	(0.017)***	(0.017)***	(0.017)	(0.017)
Constant	32.044	34.514	39.017	40.302	32.994	36.600
	(3.249)***	(3.384)***	(3.889)***	(4.064)***	(3.908)***	(4.095)***
Observations	41535	41532	17280	17279	24255	24253
R-squared	0.17	0.17	0.08	0.08	0.08	0.08

* see note Table 4.3.

Table A6

*Robustness checks using the endogenous share of population with higher education**

Dependent variable:	Test scores in English					
Sample:	All students		Students with highly educated parents		Students with low educated parents	
	OLS	IV	OLS	IV	OLS	IV
The share of students in private schools	-8.573	-2.652	-11.877	-3.091	-8.697	-4.506
	(5.231)	(3.882)	(6.016)**	(5.271)	(5.503)	(4.050)
Inverse Mills Ratio	-1.372	-1.317	-0.499	-0.522	-0.284	-0.077
	(1.440)	(1.457)	(1.628)	(1.648)	(1.500)	(1.507)
Herfindahl Index of public school competitiveness in municipality	0.377	-0.005	0.877	0.451	-0.081	-0.770
	(0.817)	(1.017)	(1.185)	(1.429)	(0.822)	(1.067)
The share of population with higher education	-1.311	-1.256	0.993	0.804	1.394	2.079
	(2.583)	(2.837)	(3.089)	(3.480)	(2.751)	(3.007)
Herfindahl Index of educational level in municipality	-6.665	-6.618	-2.205	-1.919	0.412	-0.198
	(8.044)	(8.232)	(10.633)	(11.032)	(8.585)	(8.715)
The share of employees between 20 and 54 municipality	5.238	4.219	5.298	3.890	4.072	2.113
	(3.164)*	(3.874)	(3.892)	(4.908)	(3.448)	(4.161)
The share of immigrants in municipality	10.986	10.629	14.632	14.056	10.953	10.775
	(4.825)**	(4.758)**	(6.096)**	(6.008)**	(5.160)**	(5.093)**
Population size (in 1000)	0.000	-0.000	-0.001	-0.002	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
The school cost per student in municipality	0.058	0.053	0.055	0.045	0.065	0.062
	(0.018)***	(0.018)***	(0.025)**	(0.024)*	(0.018)***	(0.018)***
Municipal Revenue	-0.044	-0.041	-0.069	-0.061	-0.037	-0.035
	(0.020)**	(0.020)**	(0.027)**	(0.027)**	(0.019)*	(0.019)*
Constant	37.085	37.240	45.276	45.618	36.598	37.244
	(4.085)***	(4.329)***	(5.382)***	(5.759)***	(4.260)***	(4.490)***
Observations	41023	41020	17252	17251	23771	23769
R-squared	0.14	0.14	0.05	0.05	0.06	0.06

* see note Table 4.3.

Chapter 5
Gender Gaps in Student Achievement: Evaluation Schemes and Teacher Characteristics

Gender Gaps in Student Achievement: Evaluation Schemes and Teacher Characteristics

Linn Renée Naper*

Abstract

This paper compares student achievement measures from three different evaluation schemes in Norway. First, using difference-in-difference estimation, the analysis reveals substantial gender biases in teacher assessment. Conditional on individual test scores from anonymously evaluated central exams, girls receive significantly higher marks than boys when assessed by the teacher. The bias is present in both mathematics and languages (Norwegian and English). Exploiting intertemporal variation within schools, teachers are found to assess same-sex students less favourably than opposite-sex students in Norwegian language classes. Hence, the observed gender bias seems to be related to teacher grading practices. Second, comparing individual student performance from a high-stake central exam and a low-stake national test, girls are found to perform significantly better when stakes are high. Girls' relative ability to perform when exposed to a high-stake one-day test is therefore not a likely explanation for the observed gender bias in teacher assessment.

Keywords: evaluation schemes, teacher assessment, student-teacher gender interactions
JEL-classifications: I21

1. Introduction

Student achievement in terms of marks and test scores is frequently used to measure educational production and performance, and individual achievement levels on different tests are important in applications for higher education as well as for future job prospects. It is therefore important for the authorities in charge of public education systems to ensure that student achievement is evaluated and measured as objectively and correctly as possible. For enrolment in non-compulsory education, colleges in the US mainly rely on the SAT undertaken after finishing high school, while many European countries rely on marks set by teachers and based on assessment during the school year[1].

* Helpful comments from Torberg Falch and Lars-Erik Borge are gratefully acknowledged.
[1] http://www.eurydice.org/portal/page/portal/Eurydice/Overview/OverviewByCountry

In addition, several countries have a high-stake central exam where each student has to participate[2]. The present paper compares the outcomes from different testing and evaluation schemes using information from Norwegian lower secondary schools. Motivated by the general concern about gender discrimination and the observed gender gap in international students' assessment surveys, such as PISA and PIRLS (e.g. OECD, 2004 and IEA, 2004), I focus on estimating gender bias effects in teacher assessment. Moreover, exploiting individual student achievement levels from three different evaluation schemes and information about teachers, I investigate how the estimated gender biases in teacher assessment are related to student behaviour and potential gender interactions among students and teachers.

Student evaluation schemes generally differ along three dimensions. First, evaluation may be anonymous or non-anonymous. Second, evaluation may be based on student performance on a single day or over a longer period. Third, evaluation may have important implications for individual students' prospects for admission to higher education (high-stake), or it may just serve as an instrument for comparison of schools with no implications for the individuals (low -stake)[3]. Empirical evidence on the heterogeneity across different types of evaluation schemes is important to consider how well the different testing schemes measure educational performance. Existing evidence suggests that different methods of assessment are not gender neutral (Powney, 1996), and that a higher weighting of course work elements improves the relative performance of girls (Stobart et al., 1992). According to Borghans et al. (2006), individual effort and achievement may also depend on the reward related to the result. In this respect, there is a concern related to the reliability of low-stake tests. Results from various surveys (like PISA, PIRLS, and TIMMS) are widely used in the economics of education literature, and the empirical evidence indicates that such tests do not measure the same as, for example, central high-stake exit exams. Several papers emphasize that girls tend to outperform boys when it comes to cognitive tests and evaluations, and according to Micklewright and Schnepf (2004), there is a systematic gender achievement gap in favour of girls on international student assessment surveys. Educational research also shows that female students outperform male students (see Ding et al., 2007 for a contribution from the US). A Swedish contribution (Emanuelsson and Fischbein, 1986) shows that girls, given the

[2] Several papers, including Bishop (1998), Wößmann (2003) who examines international data sets and Jürges et al. (2005) who study variation within Germany, find that the central exit examination is positively related to students' performance.
[3] Test results may also involve economic incentives for the schools and school owners, giving the teachers and schools an incentive to raise student performance levels. Some papers have evaluated the reliability of tests used in accountability systems. See, for example, Jacob, (2007), Jacob and Levitt (2003) and Kane and Staiger (2002).

achievement level in test scores, receive comparatively higher marks than boys when assessed by the teacher.

The observed gender gap is often explained by pointing to the general increase in the female teacher share. Based on data from the US and England, Ammermueller and Dolton (2006) and Dee (2005a, 2005b), among others, present evidence that students profit from having a same-sex teacher. Lindahl (2007) reports positive same-sex teacher effects for Swedish lower secondary students on national test results, while Holmlund and Sund (2006) find no strong support for the hypothesis of a positive same-sex teacher effect in their analysis of teacher assessment in Swedish upper secondary schools. Steel (1997) discusses a phenomenon referred to as “stereotype threats” as an explanation of how demographic matches (for example, with respect to gender, race or ethnicity) between students and teachers may influence educational outcomes. The idea is that students’ academic self-confidence, and therefore their performance, is limited by possible and perceived stereotypes in the classroom. For example, female students with male teachers may be staged academically by the mere perceived possibility that male teachers may discriminate against them, although they may not necessarily do that. Another potential explanation, often referred to as “role-model” effects, is that the presence of a demographically similar teacher may raise students’ academic motivation and expectations, and thus positively affect performance.

Both “stereotype threats” and “role-model” effects are so-called “passive teacher effects” and cannot be related to the behaviour of teachers. Moreover, “passive teacher effects” do not add to the understanding of systematic differences in performance across testing and evaluation schemes[4]. A concern in this respect is related to anonymous versus non-anonymous evaluation. The question is whether we can rely on the objectivity of teachers compared with an external examiner. With non-anonymous evaluation by the teacher, discrimination of certain student subgroups through “active teacher effects” is a possibility. The teacher’s expectations of the relative performance of students may be biased by general stereotype beliefs, and teacher assessment may unintentionally reflect the expectations of a student’s cognitive skills. Teachers may also more actively favour students who are most similar to themselves, something that would work in the same direction as a “role-model” effect. Also, teachers may respond to established perceptions of discrimination by compensating students

[4] One might argue that a “stereotype threat” is stronger when assessed by the teacher than when taking an anonymous exam. However, given that the student is actually staged academically by the threat of being discriminated, it is not obvious that individual performance should increase on an anonymous test relative to day-to-day performance in class.

that are most likely (expected) to face discrimination by assessing them relatively highly[5]. This is a kind of "active teacher-effect" that potentially would work in the opposite direction to that of "stereotype threat" effects.

Lavy (2004) tests for the existence of gender stereotyping and discrimination by public high-school teachers in Israel, using student achievement from a blind and a non-blind test available in nine different subjects. Contrary to expectations, he finds that the gender bias is against male students, meaning that male students, and not female, face discrimination in each of the subjects tested. Moreover, in his paper, Lavy argues that the observed negative male "bias" in non-blind scores may result from students not being at the same intrinsic relative ability level when taking the two different tests. The blind test is taken later than the non-blind test and on average, girls may prepare earlier than boys and, therefore, perform relatively better on the non-blind test. However, the empirical results presented by Lavy show that the discrimination against male students is related to male teachers, and that the gender bias is not likely to result from teachers adjusting the scores in response to individual student behaviour. Lavy concludes that the bias in test scores in favour of girls is the result of teachers' and not students' behaviour. A related study by Lindahl (2007) on Swedish lower secondary education finds conditional on test scores that same-gender teachers are less generous than opposite-gender teachers when assessing students' performance in mathematics, but not in languages (English and Swedish)[6].

The present paper exploits information about individual student achievement in mathematics, Norwegian and English from Norwegian lower secondary schools. In the last year of compulsory education, students are assessed by their teacher (teacher assessment) and they undertake a written central exam. Data are available for four subsequent years and reveal that, on average, girls outperform boys in all subjects, both regarding teacher assessment and central exam results. The marks from both evaluations schemes matter for upper secondary school admission prospects and are regarded as high-stake tests. However, the two achievement measures differ in the sense that teacher assessment is non-anonymous and is based on student performance during the whole school year, while the central exam is an anonymously evaluated one-day test taken at the end of the school year.

[5] Male teachers may be more concerned about discrimination towards girls than female teachers, and therefore tend to give girls an upward adjustment in assessment to compensate for the likelihood of being discriminated.
[6] In a study by Lindahl (2007) the negative same-gender teacher effect found on teacher assessment offsets the reported positive same-gender effect on test scores.

First, I compare teacher assessment and central exam results across gender in a difference-in-difference analysis. The focus is not whether girls obtain higher or lower marks than boys, but rather whether the difference between teacher assessment and exam result is significantly different across the genders. Thus, a gender bias is observed if teacher assessment is significantly different across gender, conditional on the observed exam results. The estimation results illustrate that in all three subjects, there are substantial gender biases in teacher assessment. The estimated effects are positive and significant, indicating that the bias in teacher assessment is in favour of girls. Thus, the difference between teacher assessment and exam results is significantly larger positive for girls than for boys. The estimated gender effect is smaller in Norwegian than in mathematics and English.

The empirical results reveal that the two evaluation schemes produce different outcomes across gender. Still, we cannot *a priori* say whether the observed gender bias in teacher assessment is a result of differences in evaluation strategy (anonymous vs. non-anonymous) or testing strategy (one-day test vs. assessment over time). In the second part of the analysis, I therefore apply different strategies to approach this problem. First, if the gender bias is due to differences in evaluation strategies, we expect the bias to be correlated with characteristics of the teachers. By relating the estimated gender bias to information about the gender of the teachers, I can test the hypothesis of a student–teacher gender interaction effect. Although I cannot test explicitly whether gender bias effects are due to “active” or “passive” teacher effects, passive teacher effects are expected to exercise the same effect on student achievement regardless of evaluation strategy. Thus, a significant student–teacher gender interaction effect in this setup would indicate that teachers actively adjust teacher assessment grades depending on students’ gender. Grading policies may differ systematically between schools, and the estimated gender bias and interaction effects may therefore be biased if students and/or teachers are systematically sorted between schools. I approach this potential problem by including school fixed effects in the empirical analysis. Individual students cannot be linked to their classroom teacher and, with school fixed effects, the identification thus rests on intertemporal variation in teacher composition within schools. In accordance with the results in Lavy (2004) and Lindahl (2007), the empirical results indicate that the estimated gender gap in Norwegian is related to the gender of the teacher and, moreover, that male teachers on average tend to assess girls more favourably than female teachers do[7].

Teacher assessment and central exam results are equally weighted in students' grade point average, and differences in rewards are not a likely explanation for the observed gender bias. Nevertheless, although the estimated student–teacher gender interaction effect indicates that the gender bias is related to the evaluation strategy, at least in Norwegian, I cannot rule out that the estimated gender bias is also related to differences in testing strategies (evaluation over time vs. one-day testing). A potential hypothesis is that girls and boys handle high-stake, one-day tests (such as the central exam) differently, and that this is reflected by the observed gender bias in teacher assessment conditional on individual exam results. If girls handle one-day, high-stake tests worse than boys due to nerves, for example, girls are likely to perform better on a one-day test when the stakes are low[8]. This could potentially explain why girls obtain relatively better results when assessed by the teacher, as the evaluation is then not based on performance on a single day. Exploiting individual student achievement from an anonymously evaluated national low-stake test, I can identify gender bias effects in high-stake exam results relative to results from the low-stake national test using the same difference-in-difference strategy as earlier. Contrary to expectations, the empirical results show that girls perform significantly better on the high-stake central exam than they do on the low-stake national test. Thus, the estimation results do not support the hypothesis that the gender bias in teacher assessment may result from girls' general ability to perform on high-stake tests. The findings indicate that the gender bias in teacher assessment is related to differences in evaluation strategies and moreover, provide support for the finding that male teachers assess female students more favourably than they assess male students.

2. Institutional Setting

Primary and secondary education in Norway is mainly publicly provided. The municipalities operate primary and lower secondary schools, and the counties run upper secondary schools. There is a legal right to 13 years of schooling. The first 10 years are compulsory (primary and lower secondary) and students are assigned schools according to a neighbourhood rule in the municipality they reside in. In 2005, 1164 local public schools in Norway provided education

---

[7] More generally, the results can be interpreted as an opposite-sex teacher effect as in Lindahl (2007), meaning that teachers assess students of the same sex as themselves relatively tougher than students of the opposite sex.

[8] Gneezy et al. (2003) suggest that women can be less effective than men in competitive environments, even if they are able to perform equally well in non-competitive environments. There are however also contradictory evidences on this topic, see for instance Paserman (2007).

at the lower secondary level, 662 schools were mixed schools with students at all levels (1st to 10th grade) while the remaining 502 schools only had students at the 8th to 10th level.

At the lower secondary level (8th to 10th grade), students are evaluated both non-anonymously by their teachers (marks given in all curricula-based subjects) and anonymously in central exams. Teacher assessment is based on class work throughout the school year. Central exams take place at the end of the final year (10th grade). Teacher assessment and central exam results are equally important for students' final grade point average at the end of compulsory school, and thus for prospects of admission to upper secondary school. Since the same reward is related to the result, both teacher assessment and central exam results are regarded as high-stake tests. In addition to these two high-stake achievement measures, during spring 2004 a third student evaluation scheme was introduced in form of a national test taken during the last semester of the 10th year. These tests are introduced as instruments to evaluate and monitor schools and provide feedback to teachers, students and school owners throughout the school year[9]. An external evaluation group anonymously evaluates the national tests, but the results are not included in students' final grade point average and have no consequences for entrance to upper secondary schools. The national tests are therefore regarded as low-stake tests.

According to the Educational Act of 1992[10], the main purpose of evaluation and grading during the school year is to encourage students' learning and progress. About 20 per cent of the students are examined in Norwegian and roughly 40 per cent are examined in mathematics and English[11]. In addition to the written exam, students may also take an oral exam. Class work and external exams are evaluated using a scale from 1 to 6, where 6 is best. At the end of the 10th year, all students receive a compulsory school diploma with teacher assessment in all subjects and the written and oral exam results. Students apply for an additional three years of upper secondary education based on their final grade point average.

The Norwegian Directorate for Education and Training prepares the written central exams[12], while local authorities are responsible for the assignment of examination subjects to schools and individual students. The teachers or schools have no influence in this respect. The students' responses are corrected and evaluated by two external examiners according to

[9] The national tests have only been completed twice (spring 2004 and 2005) and after change of government in 2005, the decision was made to revaluate and go through the testing system.
[10] See http://www.lovdata.no/ for the Educational Act and instructions.
[11] Students may be exempted from the external exam because of illness on examination day, disabilities, etc. The written exam absence rate is close to 4 per cent each year.
[12] Utdanningsdirektoratet.

guidelines provided by the Directorate. Teacher assessment is the responsibility of individual teachers, but the law also provides clear guidelines concerning this process. Teacher assessment is based on students' effort and achievement throughout the school year, and should express the students' competence and skills relative to the expected level of the given grade. The students' initial level of skills and qualifications should not influence or be reflected in the given marks. In subjects where students also undertake a central exam, teacher assessment should be given at least one day before the notification of examination results. Thus, teachers always assess their students independently of the exam results.

3. Data and Descriptive Statistics

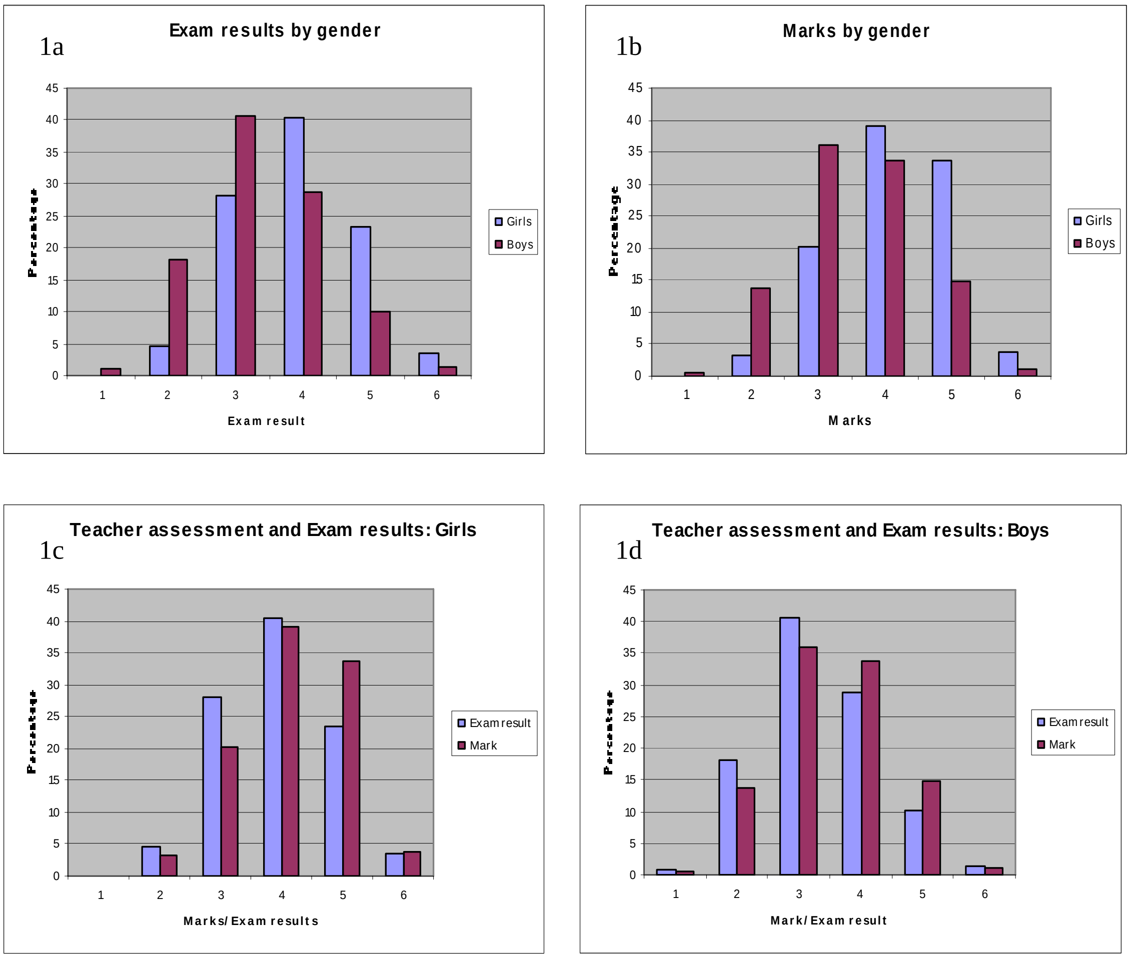
Information on students and teachers in lower secondary schools is provided by Statistics Norway. Information on individual student achievement from teacher assessment, central exams and national tests[13] from the school years 2001/2002 to 2004/2005 is merged with extensive information on individual student background, such as gender and immigration status and parents' income, marital status and education[14]. Information about individual teachers includes teachers' gender, teaching experience, marital status and number of children. The teacher information is aggregated at the school level using individual work-time percentages, and merged with student level data using a school code.

I focus on within-student variation in achievement depending on evaluation scheme. The estimation sample only includes students with both teacher assessment and central exam result in a given subject. The data do not indicate whether teachers work in primary or lower secondary levels and, to avoid linking primary school teachers to students at lower secondary levels, I restrict the analysis to pure lower secondary schools (only 8^{th} to 10^{th} grade). By reducing the estimation sample to pure lower secondary schools, I include roughly 76 per cent of the total student population. Including only students registered with both teacher assessment and central exam result in each subject, the final estimation samples for marks over the observed four years is 55 517 student observations in mathematics, 52 113 in English and 31,415 in Norwegian. The estimation samples for national tests, where only one year of data is included, have 10,117 observations in mathematics, 12 103 in English and 6 493 in Norwegian.

[13] Note that national test results are only available for students in 2003/2004.

[14] For the two first years there is also information on age (date of birth), number of siblings and birth order, parents' labour market status and sector. Including these covariates does not change the empirical outcomes of the analysis in section 5, but in order to exploit the year-to-year variation, I only use individual characteristics available in all four years.

Figure 1a–d
Student achievement in Norway: distributions by gender and evaluation type



Figures 1a-1d illustrate the gender distributions of teacher assessment and central exam results in Norwegian. The distribution of exam results in 1a is close to normal and with an overweight of girls in the upper part. Comparing 1a with 1b, we clearly see that the distribution of teacher assessment is somewhat skewed to the right and with more girls in the upper part. In 1c and 1d, I present the distribution of individual teacher assessment and exam results for girls and boys, respectively. Fifty to sixty per cent of the students are assessed equally by their teachers and in the central exam. However, the share of students with poorer exam results than their teacher assessment is higher for girls than for boys. This pattern is similar for English and mathematics.

Table 1a compares mean teacher assessment and central exam results across genders for mathematics, English and Norwegian. For each subject the table report mean achievement level and the test-statistic from a mean comparison test across gender and evaluation schemes.

The average achievement is higher for girls than for boys in all cases and all test statistics presented in the last column of the table are clearly significant at conventional levels.

Table 1a
Mean comparison tests by gender: national test scores and exam results

Subject	Mathematics			English			Norwegian		
	Test score	Exam result	t-value*	Test score	Exam result	t-value	Test score	Exam result	t-value
All									
Mean (St dev)	3.28 (1.11)	3.28 (1.11)	0.00	3.66 (0.97)	3.66 (0.97)	0.00	3.69 (0.97)	3.69 (0.97)	0.00
Girls									
Mean (St dev)	3.23 (1.08)	3.29 (1.09)	–2.46	3.83 (0.99)	3.81 (1.00)	0.84	3.85 (0.98)	4.02 (0.89)	–7.51
Boys									
Mean (St dev)	3.32 (1.14)	3.27 (1.13)	2.30	3.49 (1.08)	3.50 (1.07)	0.77	3.54 (1.02)	3.38 (0.94)	6.69
t-value	–3.96	0.79		18.01	16.37		12.82	27.84	

The gender gaps are largest in Norwegian and smallest in mathematics. The presented t-values also indicate that the gender gaps are larger for teacher assessment than for central exam results. The t-values reported in rows (for mathematics 33.42 overall, and 26.58 and 20.92 for girls and boys, respectively) show that average student achievement is significantly higher when students are assessed by their teacher than on external exams. Note also that the achievement gap for girls is quantitatively larger than for boys (0.25 vs. 0.20 in mathematics).

Table 1b
Mean comparison tests by gender: marks and exam results

Subject	Mathematics			English			Norwegian		
	Mark	Exam result	t-value*	Mark	Exam result	t-value	Mark	Exam result	t-value
All									
Mean (St dev)	3.45 (1.14)	3.22 (1.15)	33.42	3.74 (1.07)	3.57 (1.08)	24.99	3.82 (0.98)	3.62 (0.98)	25.08
Girls									
Mean (St dev)	3.51 (1.12)	3.26 (1.13)	26.58	3.96 (1.01)	3.76 (1.02)	22.19	4.13 (0.89)	3.92 (0.92)	19.89
Boys									
Mean (St dev)	3.39 (1.15)	3.19 (1.16)	20.92	3.52 (1.08)	3.39 (1.09)	14.15	3.52 (0.96)	3.34 (0.96)	17.41
t-value	12.44	6.70		48.39	40.90		57.43	55.09	

Table 1b compares mean national test scores and central exam results for students in the school year 2003–2004 and presents the same mean comparison tests as in Table 1a. Because national tests are evaluated using a different grading scale than for teacher assessment and exams, I impose the same distribution (mean and standard deviation) on the national tests as the exam results. The test statistic from comparing test scores and exam results for all students is thus zero. The results from comparing means across gender reveal that girls perform relatively better on high-stake central exams while boys perform relatively better on the low-stake national test[15]. These differences are significant in mathematics and Norwegian. In English, there are no clear differences. Girls outperform boys on the national test in Norwegian and English, while boys do slightly better on the national test in mathematics. Note also that for this sample of students, there is no significant gender gap in exam results for mathematics.

Table 2
Teacher characteristics: aggregated at school level

Teacher characteristic	Mean (St dev)
Share of female teachers	0.54 (0.11)
Mean experience of teachers (years)	19.96 (3.50)
Share of teachers with no children	0.18 (0.11)
Share of married teachers	0.64 (0.12)

In Table 2, I report the available school level characteristics of teachers. There are slightly more female than male teachers employed at the schools[16]. The mean experience level for lower secondary teachers is about 20 years, 64 per cent are married and 18 per cent have no children.

[15] As the national test scores originally are assessed using another scale than the central exams, it is important to emphasize that figures concerning these tests only are interpreted in terms of *relative* performance and not in terms of test score *levels*.
[16] For mixed schools, the shares are 65 per cent female and 35 per cent male teachers; thus, the male teacher share is considerably lower at the primary levels (1st to 7th grade).

4. Econometric Specification

4.1 Estimating the gender bias in teacher assessment

With access to two different achievement measures for each student, I can estimate the difference between teacher assessment and exam results across genders using the following linear estimation model[17]:

$$(1) \qquad A_{eijt} = \alpha + \lambda G_{ijt} + \delta E_{eijt} + \gamma \left(E_{eijt} \text{ x } g_{ijt} \right) + \mu_t + \sigma_{eijt}$$

In equation (1) individual achievement A_{eijt} (e=1: teacher assessment, e=0: central exam results) of student *i* in school *j* at time *t* is assumed to be a function of student gender G and the type of evaluation E. Evaluation is non-anonymous (E=1) for teacher assessment and anonymous (E=0) for central exam results. Each student is observed at one point in time, at the end of 10th grade, and fixed years effects μ_t are included to control for the year the individual and the school is observed[18]. The parameter σ_{eijt} is an assumed white noise error term. The model is estimated separately for each subject.

Using a difference-in-difference strategy as in equation (1), all individual and school fixed effects are implicitly assumed away with regard to the parameter γ, as long as these effects are homogenous across evaluation schemes (Lavy, 2004). The difference-in-difference parameter γ identifies the mean gender difference in achievement gaps[19]. A positive and significant γ would indicate that girls, conditional on the individual exam results, receive higher marks from their teachers than boys. The parameters λ and δ identify the gender difference in teacher assessment and the achievement gap for boys, respectively[20]. With reference to the methodology on difference-in-difference estimation, consistency of the difference-in-difference parameter occurs if assignment to the treatment is random. In this case, this means that the gender bias γ is consistently estimated as long as girls are not systematically sorted into schools where teachers evaluate girls more generously. In other

[17] As the grading scale is ordinal ranging from 1 to 6, an ordered multinomial model would be appropriate. However, the average gender difference of interest is more easily obtained from a linear model, and for simplicity linear estimates are presented.

[18] School policies, organization, the appointed principal, teachers' grading practice, etc. is likely to affect student achievement, and may also affect final grades and examination results differently. With fixed year effects, we do not have to bother about changes in such factors over time.

[19] Estimating γ from (1) is algebraically the same as estimating it from the following equation: $A_{e=1,ijt} - A_{e=0,ijt} = \Delta A_{ijt} = \delta + \gamma G_{ijt} + \Delta\sigma_{ijt}$

[20] The achievement gap for girls is thus defined as $\delta + \gamma$.

words, consistency of γ is ensured if the assignment of girls to schools is not systematically related to teacher grading practice, and not systematically different from sorting of boys.

If schools are not homogenous with respect to grading practice, the estimated gender bias in teacher assessment (the parameter γ) is likely to be biased. Schools with a high share of students from families with low socioeconomic status and subsequent 'poor reputation' may, for example, systematically exercise a more generous grading practice than other schools. Also, schools with a poor reputation may on average attract less qualified teachers. In the empirical analysis in section 5, the model is therefore extended with school fixed effects. School fixed effects as a control for variation in teacher grading practices and quality across schools is particularly important if grading practice and teacher quality are correlated with the gender of the students.

Individual student background characteristics are important determinants for student achievement (Hanushek, 2002; Hægeland et al., 2004). Existing literature also indicates that individual teacher characteristics, such as experience levels and education, may exercise an important impact on student achievement levels (e.g., Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al., 2007). If the effects of individual student and school/teacher characteristics are not homogenous across evaluation schemes, it is important to control for such factors to increase the precision of the estimates. Therefore, the model is also extended by including a large set of student and teacher characteristics. In particular, I control for parental education, income and marital status and students' migration status [21], and observable teacher characteristics as presented in Table 2.

4.2 Explaining the gender bias in teacher assessment

As discussed in the introduction, we cannot *a priori* say whether an observed gender bias in teacher assessment, for given exam results, is a result of differences in evaluation strategy (anonymous vs. non-anonymous) or testing strategy (one-day testing vs. assessment over time). Assuming that teachers do not influence students' ability to handle one-day testing, I first attempt to distinguish between these potential explanations by investigating how the gender bias interacts with teacher characteristics. By relating the estimated gender bias to information about the gender of the teachers, I can test for potential student–teacher gender

[21] Data also provide information on ethnic background (country of birth), number of siblings, birth order and birth month. However, this information is only available in two of four years. Excluding two years of data, and estimating the gender (social background) effects using this additional student level information, does not alter the empirical results in section 6.

interactions. Second, I test the hypothesis that the ability to perform on one-day, high-stake tests differs across genders by comparing individual student achievement from the high-stake central exams with performance on a one-day low-stake national test.

In equation (2), the linear model is extended by introducing a variable F, representing the share of female teachers at the school. In order to estimate the interaction between the teacher female share and the gender bias effect, the share of female teachers should be included separately, interacted with the gender–evaluation interaction $\left(E_{eijt}\, x\, G_{ijt}\right)$, with students' gender $\left(G_{ijt}\right)$ and the type of evaluation $\left(E_{eijt}\right)$ as in equation (3) below.

$$\text{(2)} \quad \begin{aligned} A_{eijt} &= \alpha + \lambda G_{ijt} + \delta E_{eijt} + \gamma\left(E_{eijt}\,\mathrm{x}\; G_{ijt}\right) + \\ &\beta_2 F_{jt} + \beta_3 (F_{jt}\, x\, E_{eijt}\, x\, G_{ijt}) + \beta_4 (F_{jt}\, x\, E_{eijt}) + \beta_5 (F_{jt}\, x\, G_{ijt}) + \varepsilon_{eijt} \end{aligned}$$

Estimating equation (2) is algebraically the same as estimating equation (3), where I rewrite the model using the difference in individual achievement ΔA_{ijt} as dependent variable.

$$\text{(3)} \quad A_{e=1,ijt} - A_{e=0,ijt} = \Delta A_{ijt} = \delta + \gamma G_{ijt} + \beta_4 F_{jt} + \beta_3\left(F_{jt}\, x\, G_{ijt}\right) + \Delta\varepsilon_{ijt}$$

The parameter β_4 provides information about mean female teacher grading practice, and a positive sign indicates that teacher assessment is higher the more female teachers at the school and vice versa. The parameter of interest β_3 indicates whether the estimated gender bias in achievement is related to the gender of the teachers. A positive sign here indicates that female teachers favour girls by raising girls' marks relative to their exam results. The parameter δ represents boys' mean achievement gap and γ the gender difference in achievement gaps[22].

As for the identification of the gender bias in equation (1), it is also important here to control for potential sorting of teachers between schools and for other school-specific characteristics that are correlated with both student achievement and student gender. By including fixed school effects in equation (3), the identification of β_3 rests on the intertemporal variation in teacher composition within schools. Additional student and teacher characteristics are included to allow for these factors to affect the gap between teacher assessment and exam

[22] The estimated γ from equation (3) is equivalent with the gender bias effect estimated in equation (1), although not quantitatively identical as the female teacher share is included in the equation.

results[23]. Passive teacher effects, as described above, are expected to exercise the same effect on student achievement regardless of testing strategy, and therefore should have no effect on the difference in achievement for individual students. Hence, an estimated student–teacher interaction effect β_3 in this setup indicates that teachers actively adjust teacher assessment depending on students' gender.

Equation (4) investigates individual student performance on low-stake tests relative to high-stake tests by replacing the dummy for evaluation strategy in equation (1) with a dummy variable *S* for "test-importance." The dummy *S* takes the value of 1 for "low-stake tests" (national tests) and 0 for "high-stake tests" (central exam).

(4) $$A_{sijt} = \alpha + \lambda G_{ijt} + \delta S_{sijt} + \gamma \left(S_{lijt} \text{ x } G_{ijt} \right) + \varepsilon_{sijt}$$

The dependent variable A_{sijt} in equation (4) is student *i*'s achievement on the central exam (s=1) and national test (s=0) in school *j* at time *t*[24]. The parameter λ represents the gender difference in central exam results and the parameter δ is the difference between central exams and national tests for boys. The parameter γ is the difference-in-difference parameter and reflects the estimated mean gender difference in relative achievement gaps. If girls handle high-stake one-day tests worse than boys because of nerves, for example, this parameter is expected to be negative. Equation (4) is also extended with fixed school effects and student and teacher characteristics, as described above.

5. Empirical Results

In this section, I present results from estimating variations of Equation (1), (3) and (4). Each equation is presented with and without fixed school effects and additional controls. In all tables, I only report the estimation results for gender, evaluation schemes and the gender-evaluation scheme interactions, while the effects of the student and teacher control variables are reported in appendix Table A1.

[23] The control variables are included similarly as in equation (1); however, included in equation (3) the estimated effects change interpretations as the dependent variable here is the deviation between teacher assessment and exam result.
[24] Student achievement on the national test is only observed in year 2003/2004; thus, year dummies are not included in this specification of the model.

5.1 Gender bias in teacher assessment

Table 3 presents the results from estimating equation (1) separately for mathematics, English and Norwegian. In all three subjects, the estimated coefficient of the interaction between the dummy for girls and non-anonymous evaluation (teacher assessment) is positive and statistically significant. The effect is smaller in Norwegian than in mathematics and English. The positive effects are in line with the results in Lavy (2004) and indicate that, conditional on individual exam results, girls are on average rewarded significantly higher by their teachers than boys are. Except for a slight reduction in the effect in Norwegian, the qualitative results are not altered by including school fixed effects and additional student and teacher characteristics. Given that the effect of student and teacher characteristic on student achievement is homogenous across different evaluation schemes, this is as expected from the difference-in-difference specification of the regression model.

Table 3
The gender bias in teacher assessment

	Mathematics		English		Norwegian	
	Eq. (1)		Eq. (1)		Eq. (1)	
Dummy for girls	0.065 (5.72)	0.059 (5.64)	0.373 (32.36)	0.375 (33.28)	0.583 (44.11)	0.600 (46.56)
Dummy for teacher assessment	0.202 (19.23)	0.198 (18.67)	0.132 (11.99)	0.131 (11.74)	0.186 (14.32)	0.192 (14.48)
Girl x (Teacher assessment)-interaction	0.055 (8.65)	0.058 (8.79)	0.064 (8.91)	0.066 (8.82)	0.020 (2.21)	0.016 (1.65)
Student characteristics	no	yes	no	yes	no	yes
Teacher characteristics	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes
Observations	111064	103113	108259	100551	62849	58222
R-squared	0.01	0.16	0.04	0.15	0.10	0.22

Note: Estimated coefficient with absolute t-value in parenthesis. The dependent variables are individual student achievement from teacher assessment and central exam. Standard errors are heteroscedasticity robust and corrected for school level clustering. All regressions include a dummy for year of observation. Estimated effects of additional student and teacher control variables are reported in the appendix Table A1.

In Table 3, we study the variation in the gap between teacher assessment and central exam results, and the size of the interaction effects should therefore be evaluated relatively to the standard deviation of the achievement gap distribution[25]. The estimated effect in mathematics is 0.058 grade points and corresponds to roughly 9 per cent of a standard deviation in the grade difference in mathematics. In English the estimated effect corresponds to 9 per cent of a standard deviation in the grade difference, whereas the effect in Norwegian is quantitatively smaller and corresponds to roughly 2 per cent of a standard deviation in the grade difference between teacher assessment and exam results.

Conditional on gender and the interaction between gender and the dummy for teacher assessment, the estimated effect of the dummy for teacher assessment in Table 3 is positive and highly significant in all subjects. Also, the dummy for girls is positive and significant in all specifications, confirming the pattern in the raw data reported in Table 1. Thus, teacher assessment is on average higher than central exam results, and girls generally outperform boys in all subjects. The effect of the dummy for girls is smaller for mathematics than for languages.

5.2 Gender bias and student–teacher gender interaction

The findings in Table 3 reflect a systematic difference in teacher assessment across genders. The next question is whether the gender biases are caused by differences in the evaluation strategies (anonymous vs. non-anonymous), by the differences in testing strategies (one-day testing vs. evaluation over time) or both. My first conjecture is that if the gender bias in teacher assessment is caused by the differences in evaluation strategies, it should be correlated with the characteristics of the teachers. Inspired by the literature on gender stereotypes and student–teacher gender interactions (e.g., Steel, 1997; Dee, 2005b; Ammermueller and Dolton, 2006), in this section I therefore investigate whether the observed gender bias in teacher assessment can be related to the gender distribution of teachers. A student–teacher gender interaction in this setup may be interpreted as a kind of teacher-initiated discrimination in assessment of students. As evaluation in languages to a larger extent allows for and involves subjective evaluation by the teacher, it may be argued that it is more reasonable to expect a form of assessment discrimination in languages than in mathematics.

Table 4 reports the results from estimating equation (3). The dependent variable is the difference between individual teacher assessment and central exam results in each subject.

[25] The achievement gap distribution is presented graphically in Figure 1. Mean achievement gap with standard errors in parenthesis in each subjects are: 0.228 (0.636) grade points in mathematics, 0.162 (0.712) in English and 0.195 (0.765) in Norwegian.

The model is extended similarly to the basic model in equation (1) by including fixed school effects and additional covariates. For each subject, the first column reports the estimated effects of the dummy variable for girls, the share of female teachers and the interaction between the two. In the second column, the estimated effects are conditioned on additional student and teacher characteristics, and school fixed effects.

Table 4
The student–teacher gender interaction effect

	Mathematics		English		Norwegian	
	Eq. (3)		Eq. (3)		Eq. (3)	
Dummy for girls	0.058	0.061	0.024	0.039	0.109	0.117
	(1.57)	(1.67)	(0.65)	(1.04)	(2.06)	(2.14)
The share of female teachers	-0.132	0.063	-0.080	0.149	0.045	-0.710
	(1.32)	(0.24)	(0.81)	(0.49)	(0.36)	(1.21)
Girl x (Share female teachers) - interaction	-0.005	-0.007	0.076	0.052	-0.163	-0.187
	(0.08)	(0.11)	(1.13)	(0.78)	(1.72)	(1.91)
Student characteristics	no	yes	no	yes	no	yes
Teacher characteristics	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes
Observations	55517	51545	54113	50264	31415	29104
R-squared	0.02	0.01	0.00	0.01	0.00	0.01

Note: Estimated coefficients with absolute t-values in parenthesis. The dependent variable is the deviation in student achievement between teacher assessment and central exam. Standard errors are heteroscedasticity robust and corrected for school level clustering. All regressions include a dummy for year of observation. Estimated effects of additional student and teacher control variables are reported in the appendix Table A1.

In all subjects, the estimated mean grading practice of female teachers is insignificant, which indicates that student performance levels on average are independent of the gender of the teacher. However, the estimated interaction effect between the share of female teachers and the dummy variable for girls is negative and statistically significant at the 10 per cent level in Norwegian. This result implies that the estimated gender gap in teacher assessment is related to the gender of the teacher, and that female teachers on average tend to assess girls more strictly than boys. More generally, teachers tend to assess same sex students more strictly than opposite sex students in Norwegian. In mathematics, the interaction effect is close to zero and statistically insignificant. In English, the qualitative effect is positive, but smaller in absolute terms compared with the effect in Norwegian, and not statistically significant.

The estimated interaction effect in Norwegian has the following quantitative interpretation. Conditional on individual exam results, an increase in the share of female teachers at the school from 0 to 1 affects the mean teacher assessment level for girls negatively with an additional 0.187 points compared with boys. This extra effect for girls corresponds to approximately 24 per cent of a standard deviation in the achievement gap distribution, which is a rather large effect. The total effect on teacher assessment of increasing the share of female teachers (from 0 to 1) is –0.897. Thus, a 10 per cent increase in the share of female teachers will reduce the average teacher assessment level for girls by 0.089 grade points. In a class with 30 students and 48 per cent girls, this corresponds to a reduction of approximately 1.3 grade points. One point (1.0) corresponds to one mark on the grading scale. Hence, on average, increasing the share of female teachers at the schools would reduce the observed gender achievement gap in Norwegian. The result that girls are assessed relatively harder than boys when assessed by female teachers in Norwegian may also partly explain why we only find weak evidence in Table 3, where we estimate the average gender bias in Norwegian irrespective of the gender distribution of teachers.

5.3 Gender bias and the ability to perform on high-stake tests

The support for teacher discrimination in Table 4 indicates that the gender bias is related to the differences in evaluation (anonymous vs. non-anonymous), at least in Norwegian. However, the bias in teacher assessment may also be related to variation in students' relative ability to perform on high-stake one-day tests (under pressure). A support for the hypothesis that girls perform relatively worse when stakes are high would indicate that the gender biases in teacher assessment also are related to differences in testing strategies.

Table 5 reports the results from comparing individual exam results with the results on the low-stake national tests as described in equation (4). The effect of the dummy for the low-stake national test is positive and statistically significant in mathematics and Norwegian, indicating that boys' performance levels on average are better on the low-stake national test compared with the high-stake central exam. However, this is not the case for English, where performance levels are not significantly different. The difference-in-difference parameter in this specification is the effect of the interaction between the dummy for girls and the low-stake national test. The estimated interaction effects are negative and statistically significant in mathematics and Norwegian, indicating that, conditional on central exam results, boys outperform girls on the low-stake national one-day tests. Thus, the hypothesis that girls perform relatively worse when stakes are high (such as the central exam) is not supported. Actually, the negative interaction effects also indicate that girls perform significantly better when stakes are high. The results for English make the picture less clear as the interaction effect here is positive, indicating that girls perform better on the low-stake national test in this subject. However, this effect is quantitatively much smaller and only marginally significant when fixed effects and additional controls are included.

Table 5
The performance on high-stake vs. low-stake tests

	Mathematics		English		Norwegian	
	Eq. (4)		Eq. (4)		Eq. (4)	
Dummy for girls	0.018	0.030	0.309	0.317	0.635	0.651
	(0.74)	(1.33)	(14.23)	(14.85)	(23.05)	(23.72)
Dummy for national test	0.052	0.048	-0.015	-0.015	0.162	0.180
	(2.23)	(2.06)	(0.58)	(0.55)	(5.90)	(6.26)
Girl x (national test score) – interaction	-0.105	-0.110	0.030	0.027	-0.330	-0.343
	(7.70)	(7.84)	(1.71)	(1.49)	(15.11)	(14.86)
Student characteristics (controls)	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes
Observations	20234	18740	24206	22528	12986	12106
R-squared	0.00	0.14	0.02	0.12	0.07	0.16

Note: Estimated coefficient with absolute t-value in parenthesis. The dependent variable is individual student achievement on low-stake national test and central exam from year 2003/2004. Standard errors are heteroscedasticity robust and corrected for school level clustering. Estimated effects of additional student and teacher control variables are reported in the appendix Table A1.

Based on the findings in Table 5, there is no support for the hypothesis that girls' performance is relatively worse when exposed to the high-stake central exam. Thus, the empirical results do not indicate that gender differences in teacher assessment may be explained by the differences in testing strategies.

6. Concluding Remarks

Student achievement measures are important both for individual students' admission prospect for higher education as well as for future job prospect, and as the preferred output indicators in studies of education production and efficiency. Hence, the objectivity and reliability of available performance measures are important.

Exploiting information about individual student achievement from Norwegian 10th grade students, this paper finds that, conditional on individual central exam results, girls receive significantly higher marks than boys when assessed by the teacher. Teacher assessment and central exams mainly differ with respect to the evaluation strategy (non-anonymous vs.

anonymous) and with respect to the testing strategy (evaluation based on performance over time vs. performance on a single day). These observable differences represent potential explanations for the observed gender bias. I apply two different approaches to explain the gender bias in teacher assessment. First, I test the hypothesis of a student–teacher gender interaction effect by relating the observed gender bias to the gender of teachers. In Norwegian, the analysis produces a negative and statistical significant student-gender interaction, indicating that girls on average receive significantly higher marks when assessed by a male teacher. The result indicates that the observed gender bias in teacher assessment is related to teacher behaviour and thus, to differences in evaluation strategies. Second, I test the hypothesis that girls perform relatively worse when exposed to a high-stake one-day test (resulting from nerves, for example) by comparing individual achievement levels on the high-stake central exam with achievement on a low-stake one-day national test. I find that girls on average perform better when stakes are high. Based on these results, it is therefore not likely that the observed gender bias in teacher assessment is caused by the difference in testing strategies.

This paper illustrates that prevailing evaluation schemes in Norwegian education do not produce homogenous results across student subgroups. To ensure that student evaluation is carried out as objectively and correctly as possible, future research and evaluation is needed.

References

Ammermuller, A. and Dolton, P. (2006): Pupil–Teacher Interaction Effects on Scholastic Outcomes in England and the USA, CEP, University of London, mimeo.

Bishop, J. H. (1998): The Effect of Curriculum-based External Exit Exam Systems on Student Achievement, Journal of Economic Education, 29 (2), pp. 171–182.

Borghans, L., Meijers, H. and ter Wel, B. (2006): The Role of Noncognitive Skills in Explaining Cognitive Test Scores, Research Centre for Education and the Labour Market, Maastricht University, ROA-RM-2006/5E.

Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L., (2007): How and Why Teacher credentials matter for Student Achievement, Working Paper 12828, National Bureau of Economic Research.

Dee, T. S., (2005a): A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? American Economic Review 95(2), pp. 158–165.

Dee, T. S., (2005b): Teachers and the Gender Gaps in Student Achievement, Working Paper No. W11660, National Bureau of Economic Research.

Ding, C. S., Song, K. and Richardson, L. I., (2007): Do Mathematical Gender Differences Continue? A Longitudinal Study of Gender Difference and Excellence in Mathematics Performance in the U.S., Educational Studies: Journal of the American Educational Studies Association, 40 (3), pp. 279–295.

Emanuelsson, I. and Fischbein, S. (1986): Vive la Difference? A Study on Sex and Schooling, Scandinavian Journal of Educational Research, 30 (2), pp. 71–84.

Gneezy, U., Niederle, M. and Rustichini, A., (2003): Performance in Competitive Environments: Gender Differences, Quarterly Journal of Economics, 118(3), pp. 1049-1074.

Hanushek, E. (2002): Publicly Provided Education, Working Paper 8799, National Bureau of Economic Research.

Holmlund, H. and Sund, K., (2006): Is the Gender Gap in School Performance Affected by the Sex of the Teacher?, Swedish Institute for Social Research, forthcoming in Labour Economics.

Hægeland, T., Raaum, O. and Salvanes, K.G. (2004): Pupil Achievement, School Resources and Family Background, Mimeo, Ragnar Frisch Centre for Economic Research and Statistics Norway.

Martin, M. O, Mullis, I. V. S. and Gonzales, E. J. (2004): International Achievement in the Processes of Reading Comprehension, International Association for the Evaluation of Educational Achievement (IEA), International Study Center, Lynch School of Education, Boston College.

Jacob, B. A. (2007): Test-based Accountability and Student Achievement: an Investigation of Differential Performance on NAEP and State Assessment, Working Paper 12817, National Bureau of Economic Research.

Jacob, B. A. and Levitt, S. D. (2003): Rotten Apples: and Investigation of the Prevalence and Predictors of Teacher Cheating, The Quarterly Journal of Economics, 118 (3), pp. 843–877.

Jürges, H., Büchel, F. and Schneider, K., (2005): The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany, Journal of the European Economic Association, 3(5), pp. 1134–1155.

Kane, T. J. and Staiger, D. O., (2002): The Promise and Pitfalls of Using Imprecise School Accountability Measures, Journal of Economic Perspectives, American Economic Association, vol. 16(4), pp. 91-114.

Lavy, V. (2004): Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment. Working Paper 10678, National Bureau of Economic Research.

Lindahl, E., (2007): Gender and Ethnic Interactions among Teachers and Students – evidence from Sweden, Institute for Labour Market Policy Evaluation, Working Paper 2007:25.

Micklewright, J. and Schnepf, S. V. (2004): Educational Achievement in English-Speaking Countries: Do different Surveys Tell the Same Story?

OECD (2004): Learning for Tomorrow's World – First Results from PISA 2003 Paris: Organization for Economic Co-operation and Development.

Paserman, D. M. (2007): Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players, IZA Discussion Paper No. 2834 –June 2007.

Powney, J. (1996): Gender and Attainment – A Review, Edinburgh, Scottish Council for Research in Education.

Rivkin, S. G., Hanushek, E. and Kain, J. F. (2005): Teachers, Schools and Academic Achievement, 73 (2), pp. 417–458.

Rockoff, J. E., (2004): The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data, AEA papers and proceedings, 94 (2), pp. 247–252.

Steel, C., M. (1997): A Threat in the Air – How Stereotypes Shape Intellectual Identity and Performance, American Psychologist, 65 (5), pp. 797–811.

Stobart, G., Elwood, J. and Quinlan, M. (1992): Gender Bias in Examination: How Equal are the Opportunities? British Educational Research Journal, 18(3), pp. 261–76.

Wößmann, L. (2003): Schooling Resources, Educational Institutions and Student Performance: The International Evidence, Oxford Bulletin of Economics and Statistics, 65 (2), pp. 117–170.

Appendix

Table A1

Complete estimation results with FE and additional controls

	Complete estimation results from Table 3			Complete estimation results from Table 4			Complete estimation results from Table 5		
	Mathematics FE	English FE	Norwegian FE	Mathematics FE	English FE	Norwegian FE	Mathematics FE	English FE	Norwegian FE
Girl	0.059	0.375	0.600	0.061	0.039	0.117	0.030	0.317	0.651
	(5.64)	(33.28)	(46.56)	(1.67)	(1.04)	(2.14)	(1.33)	(14.85)	(23.72)
Teacher assessed marks (Non-anonymous evaluation)	0.198	0.131	0.192	0.063	0.149	–0.710	0.048	–0.015	0.180
	(18.67)	(11.74)	(14.48)	(0.24)	(0.49)	(1.21)	(2.06)	(0.55)	(6.26)
Girl x (Teacher assessed marks)	0.058	0.066	0.016	–0.007	0.052	–0.187	–0.110	0.027	–0.343
	(8.79)	(8.82)	(1.65)	(0.11)	(0.78)	(1.91)	(7.84)	(1.49)	(14.86)
Student characteristics (controls)									
First generation immigrant	–0.406	–0.207	–0.258	0.049	0.003	–0.077	–0.325	–0.196	–0.363
	(10.80)	(5.79)	(6.84)	(2.35)	(0.11)	(2.23)	(4.15)	(2.40)	(3.39)
Second generation immigrant	–0.202	–0.054	–0.068	0.039	0.055	0.000	–0.176	–0.029	–0.198
	(5.13)	(1.14)	(1.49)	(2.00)	(2.19)	(0.00)	(2.96)	(0.34)	(2.10)
Student living with both parents	0.364	0.187	0.229	–0.000	0.020	0.031	0.318	0.163	0.202
	(34.16)	(20.59)	(24.17)	(0.02)	(2.36)	(2.88)	(14.25)	(8.46)	(10.14)
Higher education Father	0.497	0.415	0.362	–0.004	0.041	0.046	0.388	0.373	0.303
	(41.74)	(37.35)	(29.11)	(0.57)	(5.32)	(4.12)	(13.40)	(17.28)	(11.67)
Higher education Mother	0.453	0.385	0.353	0.035	0.043	0.043	0.406	0.326	0.286
	(32.97)	(28.28)	(28.73)	(5.98)	(5.65)	(4.21)	(14.61)	(15.66)	(9.56)
Income Father	0.000	0.001	0.001	0.000	0.000	0.000	0.003	0.001	0.001
	(1.21)	(3.58)	(3.78)	(1.13)	(1.09)	(2.03)	(6.34)	(2.33)	(2.86)
Income Mother	0.002	0.003	0.002	0.000	0.001	0.001	0.005	0.004	0.005
	(2.13)	(3.85)	(4.41)	(0.23)	(2.99)	(2.36)	(5.57)	(5.85)	(6.55)
Teacher characteristics (controls)									
Mean experience of teachers	0.003	–0.013	–0.021	0.003	–0.001	–0.038			
	(0.43)	(1.64)	(1.53)	(0.35)	(0.13)	(2.11)			
Share of female teachers	–0.116	–0.044	0.051						
	(0.56)	(0.26)	(0.12)						
Share of teachers without children	0.261	–0.104	–0.527	0.216	0.268	0.917			
	(1.37)	(0.50)	(1.56)	(0.81)	(0.99)	(2.01)			
Share of married teachers	0.150	0.288	–0.282	–0.023	0.237	0.348			
	(0.89)	(1.81)	(0.99)	(0.10)	(0.99)	(0.68)			
School fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Constant	2.512	3.025	3.517	–0.030	–0.200	0.860	2.622	3.070	2.930
	(11.22)	(13.44)	(9.17)	(0.11)	(0.56)	(1.22)	(78.14)	(105.17)	(88.72)
Observations	103113	100551	58222	51545	50264	29104	18740	22528	12106
R-squared	0.16	0.15	0.22	0.01	0.01	0.01	0.14	0.12	0.16

Note: Estimated coefficient with absolute t-value in parenthesis. The standard errors are heteroscedasticity robust and corrected for school level clustering. The lack of significant teacher-effects may be due to low time-variation in the data. Excluding the school fixed effects from the regressions makes the estimated effects more precise.