Effects of school characteristics on grades in compulsory school

Alli Klapp Lekholm

[a]Department of Education, University of Gothenburg,

Sweden

Address:

Alli Klapp Lekholm

Department of Education, University of Gothenburg

Box 300

SE 405 30 Gothenburg

Sweden

E-mail: alli.klapp.lekholm@ped.gu.se

Phone work: +46 46 855 05

Abstract

The purpose of the study was to investigate effects of different school characteristics, such as municipal and independently-operated schools, teacher experience and certification on grades in compulsory school, and the extent to which parental education confounds the relation between different school characteristics and grades. To answer these questions, multilevel, multivariate, techniques were used. The data derives from The Gothenburg Educational Longitudinal Database (GOLD), and the subjects were 99 070 9th grade students born in 1987. The analyses were performed on subject grades and scores on national tests in Swedish, English and mathematics and on school characteristics data. The results showed that school characteristics accounted for considerable amounts of variance in grades. However, when parental education was taken into account, the relations decreased considerably and, in most cases, became non-significant. The most interesting results concern the substantial confounding effect of parental education on the relations between different school characteristics and grades.

Keywords: compulsory school; family background; grades; grade assignment; national tests; school characteristics

Introduction

The effect of school characteristics on student achievement and grades is a subject of substantial interest and has been investigated by numerous researchers. Ever since the Coleman Report (1966) a large number of studies have indicated that demographic characteristics, such as family background and peer-effects, influence student achievement and grades. However, research has also revealed that factors related to schooling, such as teacher quality and teacher density, also influence achievement and grades (Darling-Hammond, 1999, 2000; Darling-Hammond, Ross & Milliken, 2006; Greenwald, Hedges, & Laine, 1996; Gustafsson & Myrberg, 2002; Hanushek, 1989, 1997).

Research has also indicated that, in addition to academic subject knowledge, grades may also reflect students' individual characteristics such as interest, effort and motivation (Brookhart, 1991; Cizek, Fitzgerald & Rachor, 1995; Cross & Frary, 1999; Klapp Lekholm & Cliffordson, in press; McMillan, Myran, & Workman, 2002). Assessments and evaluations may also differ between teachers, which, among other things, can be a cause of grade inflation and differences between schools (Cliffordson, 2004; National Agency for Education, 2007a).

Klapp Lekholm and Cliffordson (2008) investigated the dimensionality of grades in the Swedish compulsory school system and found, both on individual and school levels, subject-specific dimensions in final grades, as measured by national test scores and grades in three core subjects (Swedish, English and mathematics) as well as a common grade dimension, measured only by grades (Swedish, English and mathematics) which cuts across the grades in all three of these subjects. When the student background variable 'parental education' was related to these dimensions on the school level, the result showed that the proportion of students with high parental education levels was negatively related to the common grade

dimension, but highly positively related to the subject-specific dimensions in grades. This finding was tentatively interpreted to be due to teachers' concern with the distributive justice of grades, thus causing them to employ compensatory grading practices (Klapp Lekholm & Cliffordson, 2008).

In the current study, the main aim is to investigate the effects of different school characteristics, such as municipal and independent schools, school-size, teacher density and teacher certification, on grades in compulsory school. Another aim is to investigate the associations between parental education and the different school characteristics and their influences on grades on the school level.

Previous research

In the following section, previous research concerning different school characteristics and their influence on educational outcomes and grades is presented. Whilst, according to the research literature, there may be several different school characteristics of importance for grades, some school characteristics such as educational provision, size and location of school, and the characteristics of the teaching staff, seem to be of particular importance.

Educational provision: municipal and independently-operated schools

Most schools in Sweden are municipally-operated and most children attend a municipally-operated compulsory school. However, it is also possible to choose an independently-operated grant-maintained school. Independent schools, which must have received authorisation by the National Agency for Education, are financed by local authority grants and are open to all children. Research on the 'independent school effect' is somewhat inconsistent. Some studies have found a positive effect of independent schools on student achievement and grades,

whereas other studies have only found modest and often statistically insignificant effects of independent schools on student achievement (Somers, McEwan, & Willms, 2001). Somers et al. (2001) argued, furthermore, that many studies of 'the independent school effect' have been misinterpreted since controlling for the characteristics of student peer groups has been neglected. Indeed, Somers et al. conclude that:

> Consequently, if one defines the private school effect as the achievement differences between private and public schools net of peer group characteristics, then typical private school effects are probably biased: instead of reflecting school-based differences between private and public schools-related to resource levels, school practices or efficiency of resource use-private school effects will partly reflect the more privileged status of peer groups (p. 11).

Somers et al. argue that the relationship between independent schools and grades is positive, significant and quite large but, when controlling for the effects of student background, the achievement differences declined and, furthermore, after controlling for peer-group characteristics, the differences in achievement became even smaller, and indeed were on some occasions negative.

In Sweden only a handful of studies have investigated the effect of independent and municipally provided education on grades. Bergström and Sandström (2001) found that independent schools exert a positive influence on student achievement and grades, and that the competitive allocation of resources raises the standards in both types of schools. Likewise, Wikström (2005) found that independent schools award higher grades relative to students' performance on the Swedish Scholastic Assessment Test (SweSAT) in comparison to municipal schools, even after controlling for parental education. Wikström suggests that the result shows that independent schools award higher grades than municipal schools given the

same achievement levels. However, an alternative interpretation of this result is that student achievement was not properly measured by the SweSAT. The Swedish National Agency for Education (2007a) found that independent schools do not award higher grades, when background variables were controlled for. Gustafsson and Myrberg (2002) claimed that educational research within this area suffers from methodological problems due to omitted variables, the lack of control variables and the assumption that ordinary regression analyses are sufficient in order to account for causal relationships between different school characteristics, such as school types and grades. McEwan (2000) has also suggested that many studies investigating the independent school effect suffer from problems due to the lack of control for peer effects and socio-economic background on the school level. Studies which use an extensive control for peers, neighbourhood and school characteristics often find statistically insignificant or small independent school effects (Figlio & Stone, 1999; McEwan, 2000; Somers at al., 2001).

School size and school location

Darling-Hammond et al. (2006) reviewed a large number of studies and found that smaller schools seem to produce favourable student outcomes, such as higher achievement and lower drop-out rates. However, there are alternative interpretations of this relation since the influence of school size on student outcomes and grades may be confounded by other factors such as school organization, and academic and social factors (Darling-Hammond et al., 2006; Lee & Smith, 1997). Thus, schools 'small by design or 'small by default' may have different implications due to demographic features (Darling-Hammond et al., 2006; Ready & Lee, 2006). Similarly, in some areas such as for example rural areas, a certain school size may be predominant. Independent schools also tend to be smaller schools. Recent research would appear to suggest that it is not school size per se that improves student outcomes, but rather

that school size influences student outcomes indirectly through other factors such as personalization, a shared school mission, a strong academic curriculum, authentic instruction and the development of a professional community (Darling-Hammond et al., 2006; Ready & Lee, 2006).

In Sweden, Wikström (2005) found that, given the same achievement levels on the SweSAT test, there was a school size effect on grades, such that teachers in small schools (<300 students) award higher grades in comparison to second smallest (300-499 students) and the largest (>1000 students) schools, the second largest schools (500-1000) was used as the reference category. Wikström suggested that this result can be related to variation between schools in grade assignment practices, in combination with pressure for high grading, since grades function as an instrument for selection to the next level in the educational system. The pressure for higher grading on small schools may be due to the current voucher system in Sweden, which makes smaller schools more vulnerable to the loss of students.

Stanley, Comello, Edwards, and Marquart (2008) investigated the effects of rural and urban communities on school adjustment and other school-related variables. They found mediating effects of parental education, income and rurality on school adjustment. They also argued that the characteristics of students and their parents, teachers and proximate peers, were significantly related to school adjustment and perceived school performance; for example a student's achievement on standardized measurements might be due to socio-economic background, and not rurality. Other studies (Roscingo & Crowely, 2001; Williams, 2001) have also found that by controlling for socio-economic status, the relation between rurality and achievement disappears.

Several researchers draw attention to the fact that cluster effects have been neglected in many studies of school characteristics such as school size, and that the multilevel nature of the data must be accounted for in order to make correct inferences. Since school size is a school-level variable and student achievement is a student-level variable, this emphasises the need, as for example Gustafsson & Myrberg, (2002) and Lee & Smith, (1997) have pointed out, for multilevel analyses to be conducted:

> Most studies ignore this, either by aggregating achievement to the school level and running a school-level regression or by appending school size to student-level information. Both approaches can misestimate the influence of the school-level variable (in this case size) on the outcomes of interest by inadequately contending with the sources of between- and within-school variability in the dependent variable (Lee & Smith, 1997, p.175)

Teacher characteristics

Research on the influences of different teacher characteristics on student achievement and grades shows disparate results. One reason for this may be basic methodological problems such as using cross-sectional data which makes it difficult to draw causal inferences. This makes it particularly difficult to deal with the compensatory allocation of resources, such as for example that students who need more help are placed in smaller groups and are sometimes instructed by better educated teachers. Control for different selection effects (of, e.g., teachers) in the educational system is extremely important if causal inferences are to be drawn. Gustafsson and Myrberg (2002) also stress the problems with the definitions of teacher density and class size which are often used synonymously, in spite of the fact that teacher density and class size are measured differently and the effect on grades may not be the same for these two constructs.

The large amount of research on the effects of teacher density/class size on student achievement has resulted in a body of disparate findings. Glass, Cahen, Smith, and Filby (1982) concluded that there was a positive achievement effect when class size was reduced to around 15 students, but this inference has been questioned by Hanushek (1999) who argued that there is no consistent effect of class size or other resource factors. In a review of studies on the effects of class size on achievement, Gustafsson and Myrberg (2002) concluded that teacher density has an effect on student achievement but that this effect is a function of student age and background. The research findings indicate that small classes have a positive effect primarily on younger students and students with immigrant backgrounds (Hanushek, 2000; Lindahl, 2005).

The definition of teacher competence also varies greatly in the reported research. In some studies teacher competences are defined in terms of certification whereas other researchers use teacher experience or teacher effectiveness as indicators of teacher competence. These measures of teacher competence have implications for the interpretations of results when investigating how teacher competence influences student achievement. It has, for example, been suggested that compensatory resource allocation can influence results. For example, well-educated teachers might work with low-achieving students. On the other hand, experienced teachers might choose to work in schools with high-achieving students. Research on the influence of teacher competence on student performance has shown that teachers' subject competence, as well as their pedagogical competence, is of importance and, further, that these two competences interact and have a non-linear relation with student achievement (Darling-Hammond, 1999, 2000; Gustafsson & Myrberg, 2002). Darling-Hammond (1999) investigated the relations between teacher qualifications and students' achievement and found

that well qualified teachers with a good subject knowledge had students with better test results.

In a review of teacher competence and student achievement, Darling-Hammond (2000) found that, when controlling for student background, teacher preparation and certification are the strongest correlates of student performance in reading and mathematics. Further, a 'well-qualified teacher' variable showed a strong, significant relationship with student achievement after controlling for student background, which leads Darling-Hammond to suggest that this variable is a proxy for both disciplinary knowledge and knowledge of education. In a review of several studies, Gustafsson and Myrberg (2002) concluded that teacher education and experience are characteristics of importance for student achievement and that teacher competence affects student achievement to approximately the same extent as students' socio-economic background.

Research aims

The review of the research literature indicates that grades are systematically influenced by different school characteristics. However, the review also indicates that there are many methodological problems in this area of research, and that observed relations between school characteristics and grades may be due to confounding with other variables. Recently, Klapp Lekholm and Cliffordson (2008) have demonstrated that grades are multidimensional and that on the school level systematic effects of parental education on grades are evidenced.

The main aim of this study is to investigate the relations of different school characteristics and parental education on grades at the school-level. In order to answer the research questions, multivariate, multilevel techniques are used. The following questions will be addressed:

- Which school characteristics influence grades and how do these characteristics influence the subject-specific and common grade dimensions, respectively?
- To what extent does parental education confound the relations between school characteristics and the subject-specific and common grade dimensions at the school-level?

Method

Subjects

Data used in this study is derived from The Gothenburg Educational Longitudinal Database (GOLD). GOLD contains register data compiled by Statistics Sweden for all individuals born between 1972 and 1987 and where a large amount of information, such as student background, grades from compulsory and post-compulsory education, results on national tests and school characteristics is available. The subjects in the study were 99 070 students born in 1987, and who left compulsory school in 2003. In all, 1 246 schools are included in the analyses. Two reductions of the data have been made. First, individuals for whom information on both subject grades and national test results is lacking have been excluded. Secondly, schools with 14 students or fewer have been excluded, since schools with very few students are often schools for students with special needs or other special groups. In sum, 1 782 students have been excluded from the analyses.

Instruments and variables

Two measures were used, namely subject grades, and the national test scores in three core subjects: Swedish, English and mathematics. Different school characteristic variables have also been used, as well as one student background variable; parental education, which is a key variable to be controlled for. PAREDU is a dummy variable, which indicates parental

education (0=upper secondary education or lower and 1=higher than upper secondary education).

The Swedish criterion-referenced grading system

The purpose of the criterion-referenced grading system is to provide information about the individual student's acquisition of required standards, to evaluate the educational system, and to be used as an instrument for selection to the next level in the educational system. According to the National Curriculum, the purpose of grades is to measure student subject knowledge (National Agency for Education, 2004). The grading system is based on the evaluation of student attainment measured against centrally defined goals and it is highly decentralized, leaving the teachers with the full responsibility to instruct, assess and grade their students in accordance with the stipulated goals. The grading system is based on the underlying assumption that teachers will interpret the criteria in a similar way. National tests are used in order to assist teachers in calibrating their grading against the goals in the curriculum.

The grading scale used in schools consists of four levels: not pass (IG), pass (G), pass with distinction (VG) and pass with special distinction (MVG). In order to use grades for selection purposes, they are converted into numbers. The scale ranges from 0-20 where IG =0, G =10, VG =15 and MVG =20. There are no intermediate numbers. These levels reflect student attainment of the objectives or criteria for each subject. Overall standards for the final semester of compulsory school, i.e. spring of the ninth year of school, are defined centrally for all the grade levels in the curriculum. In this study, grades from the end of year nine in three core subjects will be used, namely grades for Swedish (SGSW), English (SGEN) and mathematics (SGMA).

The national tests

In Sweden, the national tests serve several purposes. They are designed to function as an instrument to elucidate the curriculum and to help the teachers calibrate their grading in order to equalise the grades between teachers and over schools. Another purpose is to evaluate student performance at school and national levels. The tests are comprised of different subtests in each subject and there are oral as well as written tests. The tests are produced centrally and the content is not revealed in advance. The national tests are curriculum-based but not all the centrally-defined goals are tested, which implies that the respective subject domains are not fully covered by the tests. In grade nine, national tests are used in three core subjects: Swedish, English and mathematics. In Swedish there are three subtests, the first test measures reading comprehension, the second is an oral test conducted in pairs and the final subtest is a written assignment. In English, the three subtests consist of oral interaction and production – usually conducted in groups – reading and listening comprehension tasks, and a short essay. In mathematics there are four subtests: an oral task done in a group, a test of arithmetic where use of a calculator is not permitted, a test with more extensive tasks, and, finally, a test which demands problem-solving and for students to account for the calculations that they make (National Agency for Education, 2007b). There are no external referees involved in the assessment and grading procedures for any of the national tests. However, teachers are strongly recommended to cooperate in the assessment and grading of the tests and the National Agency for Education has developed a test bank where teachers have access to a large amount of examples of different levels of student achievement on the national tests. The National Agency of Education (2007a) has conducted several studies which have found that the national tests are valuable instruments in achieving equivalence in grades. Studies have also shown that the national tests are reasonably valid as instruments for measuring

students' subject knowledge (National Agency for Education, 2007a; Åberg-Bengtsson & Erickson, 2006).

The scale for the national tests corresponds with the scale for the subject grades and ranges from 0-20, where not pass (IG)=0, pass (G)=10, pass with distinction (VG)=15 and pass with special distinction (MVG)=20. These are the sole scale points and no intermediate scores are used. In this paper the following abbreviations are used for the national tests: NTSW1, NTSW2 and NTSW3 for the test scores in Swedish, and NTEN1, NTEN2 and NTEN3 for English. In the case of mathematics, only one summarized test score is available, NTMA.

School characteristic variables

The school characteristic variables were collected at the end of year nine in compulsory school, in 2003. Of a large amount of possible variables, eight manifest variables on the school level have been used. Dummy variables were created to represent the provider of the schools, school size and the location of the school. The INDEPSCH variable defines schools according to the authority of the schools, (independent schools and municipal schools) where municipal schools were the reference. The SIZE variable defines schools in accordance with the number of students at the school (<300, 300-499, 500-1000 and >1000), and the smallest school size was the reference. The LOCATION variable defines the location of schools into three categories (rural schools, suburban schools and urban schools), and urban schools were taken to be the reference. Rural schools are schools located in rural areas with more than 45 minutes travel time to nearest town with 3000 or more inhabitants. The suburban schools are located in areas with 5-45 minutes travel from a town, and urban schools are located in towns and 5 minutes travel time outside the town (The Swedish National Rural Development Agency, 2008).

Measures of teacher characteristics at the school level have also been used, where TEAAGE measures the average age of the teaching staff and TEAEXP the average number of years that the teachers have worked. The TEACER measures the proportion of teachers with full certification and TEAWOM is the proportion of women. TEADENS is a measure of the number of teachers per 100 students. In this variable no administrators or other staff categories are included.

Method of analysis

In order to investigate the influence of different school characteristics on the subject-specific dimensions and the common grade dimension in grades, multilevel confirmatory factor analysis (CFA) and structural equation modelling (SEM) were used**.** The starting point has been a baseline two-level, four-factor model which identified the multidimensionality of grades on individual and school-levels (Klapp Lekholm & Cliffordson, 2008). The baseline four-factor model (A) with four latent variables (*Sw*, *En*, *Ma* and *ComGr*) was designed to reflect subject-specific dimensions (*Sw*, *En* and *Ma*) as well as a common grade dimension (*ComGr*) that cuts across the three subject grades (Swedish, English and mathematics), see Figure 1. The subject-specific factors were related to respective subject grades (SGSW, SGEN, and SGMA) and each of the national test scores (NTSW1-3, NTEN1-3, and NTMA) in Swedish, English and mathematics and with covariances between the three subject factors *Sw*, *En* and *Ma*. In order to separate the common grade dimension the *ComGr* factor was only related to the three subject grades (SGSW, SGEN and SGMA). The goodness-of-fit indices for this model were good (χ^2 (58, 99070) = 4397.08; RMSEA=.027) (Klapp Lekholm & Cliffordson, 2008). In Figure 1, the estimates on the school level are presented and they are slightly higher than the estimates on the individual level. When parental education was included in the baseline model and related to all the factors (*Sw, En, Ma, ComGr*) (model B),

the result showed that on the school level, parental education influenced the subject-specific dimensions substantially (.72, .80, .74 for *Sw*, *En*, *Ma,* respectively) and the common grade dimension negatively (-.34). Model fit was slightly better for this model (χ^2 (76, 99070) = 4786.90; RMSEA=.025).

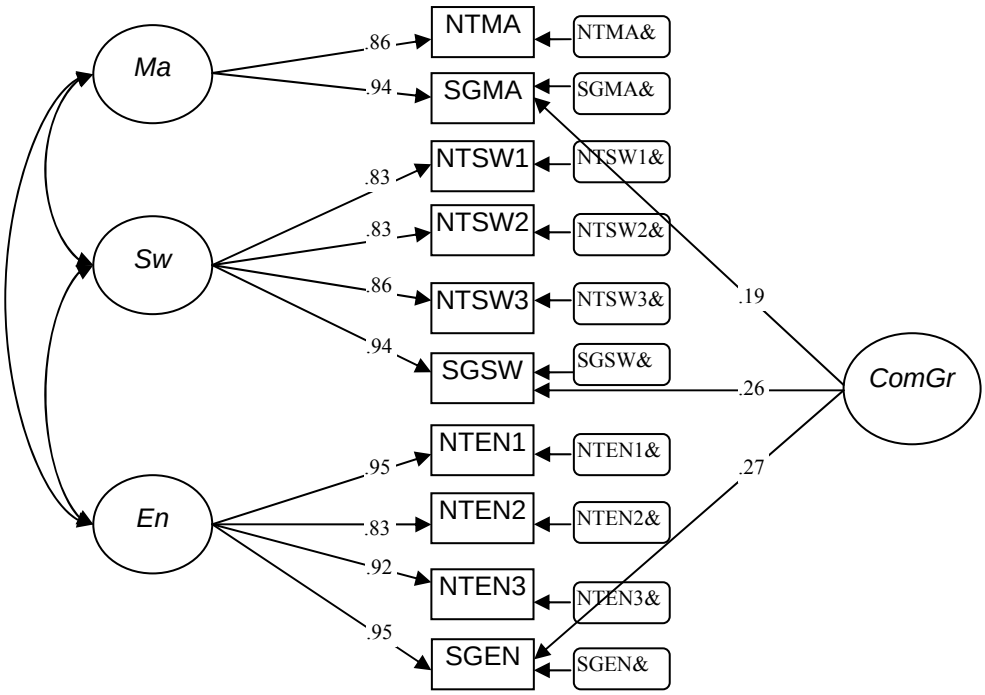

Figure 1. The baseline two-level four-factor model (A) with covariance between the subject factors, including the estimates on the school level (Klapp Lekholm & Cliffordson, 2008).

In this study, the two-level four-factor model (A) is the baseline model throughout all the analyses. When parental education is included in the models, this variable is estimated on both individual and school levels, whereas the school characteristic variables are only estimated on the school level.

The first step in this study was to estimate eight models (C1 to C8) with school characteristic variables (INDEPSCH, SIZE, LOCATION, TEAAGE, TEAEXP, TEACER, TEAWOM,

TEADENS) one at a time, and with relations to the four factors (*Sw*, *En*, *Ma* and *ComGr*) in the baseline model on the school level.

The second step was to investigate the confounding effects of parental education and school characteristics on the subject-specific dimensions and the common grade dimension in grades. Models (D1-D8) were estimated with covariance between parental education and school characteristic variables (INDEPSCH, SIZE, LOCATION, TEAAGE, TEAEXP, TEACER, TEAWOM, and TEADENS) one at a time, and with relations to the four factors *(Sw*, *En*, *Ma* and *ComGr*) in the baseline model (model A). The models also included covariances among the residuals of the three subject factors (*Sw*, *En* and *Ma*) (See Figure II).

As measures of model fit, the χ^2 goodness-of-fit test and the Root Mean Square Error of Approximation (RMSEA) were used. The RMSEA is strongly recommended as a tool when evaluating model fit since it takes both the number of observations and free parameters into account (Jöreskog, 1993). In order for a model to be acceptable the RMSEA should be below .08, while to be good, the RMSEA should be below .05. The Standardized Root Mean Square Residual (SRMR) is a measure of residuals compared separately for within and between levels. The SRMR should be below .08.

For the national tests and the school characteristics there was some missing information, which was handled with the missing data modelling procedure implemented in the Mplus program (Muthén, Kaplan & Hollis, 1987). This procedure makes the assumption that the data is 'missing at random' (MAR) which implies that it yields unbiased estimates when the missingness is random given the information in the data. This is a much less restrictive assumption than the assumption that the data is 'missing completely at random'. The fact that

there are high interrelations among the observed variables provides good possibilities to satisfy the MAR assumption (Schafer & Graham, 2002).

Mplus version 3.13 (Muthén & Muthén, 2004) was used for the purpose of estimation and for testing all the models. STREAMS (Gustafsson & Stahl, 2005) is a modelling front-end environment, which was used to execute the analyses.

Results

The means and standard deviations for subject grades, national test results, parental education and the school characteristic variables are shown in Table 1.

The descriptive statistics revealed that the missing data for most of the variables were negligible, with the exception of the national tests. For the tests in Swedish and English, the proportion of missing data ranged from 17.8 per cent for NTEN2 to 21.3 per cent for NTSW2, and in mathematics the proportion was considerable (43.1 per cent). The large amount of missing data in the national test in mathematics was due to the unfortunate fact that, in some areas in Sweden, the contents of the test had been divulged in advance. There is, however, no reason to assume that this missing information is biased due to achievement in mathematics.

Table 1. Descriptive statistics for the manifest variables; subject grades, national tests, parental educational and school characteristics.

Variables	*N*	*% missing*	*Mean*	*Std. dev*
Subject Grades				
SGSW	98 353	0.7	12.98	4.34
SGEN	98 353	0.7	13.14	4.68
SGMA	98 353	0.7	12.09	4.52
National tests				
NTSW1	81 391	17.8	11.79	4.82
NTSW2	77 979	21.3	12.91	4.00
NTSW3	81 131	18.1	12.17	4.14
NTEN1	79 832	19.4	13.15	4.25
NTEN2	81 426	17.8	13.84	5.00
NTEN3	81 221	18.0	12.73	4.40
NTMA	56 325	43.1	11.90	4.44
Student background		0.0		
Higher education[1]	44 008			
Lower education[2]	54 569			
School characteris.		0.0		
Independent school	4 311			
Municipal school	94 759			
School size		0.0		
<300	14 389			
300-499	45 878			
500-1000	38 348			
>1000)	455			
School location		0.1		
rural	1 847			
suburban	13 034			
urban	84 049			
Teacher characteris.				
Teacher age		0.1	45.5[3]	3.0
Teacher experience		0.1	14.1[3]	3.0
Teacher certification		0.1	67.2%[4]	12.3%
Teacher women		0.2	64.5%[4]	0.09%
Teacher density		0.7	9.8[5]	1.8

Note. [1] Higher than upper secondary education. [2] Upper secondary education or lower. [3] Average age and years of teaching for teachers over schools. [4] The proportion over schools. [5] The proportion of teachers per 100 students.

The baseline two-level four-factor model with school characteristics (models C1-C8)

The first step was to add the different school characteristic variables (INDEPSCH, SIZE, LOCATION, TEAAGE, TEAEXP, TEACER, TEAWOM, TEADENS), one at a time, to the baseline four-factor model (A) at the school level, and with covariances between the subject-specific factors (*Sw*, *En*, and *Ma)* at the individual level and between the residuals of the

subject-specific factors at the school level. This resulted in eight models (C1-C8) with a different school characteristic variable related to the four factors (*Sw*, *En*, *Ma*, and *ComGr*). The goodness-of-fit indices were good for all the models. Table 2 presents the regression coefficients and goodness-of-fit indices.

The relation between authority of schools and the four factors (Sw, En, Ma, ComGr) showed that the INDEPSCH (model C1) variable related strongly to the *Sw*, *En* and *Ma* factors, (.35, .42, .39, respectively), whereas no significant relation existed to the *ComGr* factor, in comparison to the reference category municipal schools.

The SIZE variable (model C2) was related with *ComGr*, which, primarily, was due to a significantly lower value on the *ComGr* factor (-.16) for the group of second largest schools (500-1000 students) than for the reference category of small schools (<300 students). The largest schools (>1000 students) had a higher level on the *Sw*, *En* and *Ma* factors (.12, .12 and .07 respectively), whilst there was no difference with respect to the *ComGr* factor in comparison to the reference category small schools.

In the model with the LOCATION variable (model C3) the result showed no significant difference between schools located in rural areas and the four factors (*Sw*, *En*, *Ma*, *ComGr*) in comparison to the reference category urban schools. However, schools located in suburban areas (5-45 minutes from a large city) had significantly lower achievement on the *Sw*, *En* and *Ma* factors (-.11, -.17, -.07, respectively), and a significantly higher level on the *ComGr* factor (.25) in comparison to the reference category urban schools.

The TEAAGE variable (model C4) and the TEAEXP variable (model C5) were negatively related to the *Sw* (-.08, -.11, respectively) and *En* (-.12, -.15, respectively) factors. The results

also showed that the TEACER variable (model C6) was significantly related to the *Ma* factor (.12). The TEAWOM variable (model C7) was positively related to the *Sw*, *En* and *Ma* factors (.16, .21, .16, respectively) and the TEADENS variable (model C8) was positively related to the *Sw*, *En* and *Ma* factors (.16, .18, .18, respectively). None of the teacher characteristic variables was significantly related to the *ComGr* factor.

Table 2. Relations between school characteristic manifest variables and the four factors (*Sw*, *En*, *Ma* and *ComGr*), on the school level (model C1-C8).

Model	Variables	*Sw*	*En*	*Ma*	*ComGr*	χ^2/df	RMSEA	SRMR W/B
C1	INDEPSCH[1]	.35	.42	.39	-.10[ns]	4411.06/64	.026	.018/.033
C2	SIZE[2]					4835.73/76	.025	.015/.029
	300-499	-.05[ns]	-.06[ns]	-.05[ns]	-.13[ns]			
	500-1000	.04[ns]	.05[ns]	.03[ns]	-.16			
	>1000	.12	.12	.07	.03[ns]			
C3	LOCATION[3]					4695.47/71	.026	.016/.033
	rural	.00[ns]	-.05[ns]	.00[ns]	.071[ns]			
	suburban	-.11	-.17	-.07	.25			
C4	TEAAGE	-.08	-.12	-.07[ns]	.03[ns]	4593.52/64	.027	.018/.035
C5	TEAEXP	-.11	-.15	-.06[ns]	.02[ns]	4495.85/64	.026	.018/.035
C6	TEACER	.07[ns]	.05[ns]	.12	-.06[ns]	4456.74/64	.026	.018/.036
C7	TEAWOM	.16	.21	.16	-.10[ns]	4488.10/64	.026	.018/.033
C8	TEADENS	.16	.18	.18	-.03[ns]	4502.72/64	.026	.018/.034

Note. [1] Municipal schools are the reference. [2] Size <300 is the reference. [3] Urban schools are the reference. All values are significant on the .001 level except for the marked values (ns) which are non-significant.

To sum up, independent schools (INDEPSCH) had a higher level of achievement on the subject-specific factors (*Sw*, *En*, *Ma*) in comparison to the reference municipal schools. The teacher characteristic variables TEAWOM and TEADENS also had positive relations to the subject-specific factors, whereas teacher age (TEAAGE) and teacher experience (TEAEXP) had negative associations to some of the subject-specific factors. Teacher certification (TEACER) had a positive relation to the *Ma* factor. The second largest schools (500-1000 students) had a higher level on the C*omGr* factor whereas the largest schools (>1000 students) had higher achievement on the subject-specific factors, in comparison to the reference category of small schools. Schools located in suburban areas had lower levels of achievement on the subject-specific factors, but a higher level on *ComGr* in comparison the reference group urban schools.

The baseline two-level four-factor model with covariance between parental education and school characteristics (Models D1-D8)

In the next step the extent to which parental education confounds the relations between the different school characteristics and the subject-specific and the common grade dimensions in grades (model D1 to D8) was investigated. Covariances were estimated between parental education and the different school characteristic variables and added to the previous estimated models (model C). Model fit was slightly better for all these models (model D1-D8). The regression coefficients, covariances and the goodness-of fit indices are presented in Table 3.

Table 3. The relations for models D1 to D8, between the different school characteristic variables, parental education and the *Sw*, *En*, *Ma* and *ComGr* factors and with covariance between the parental education variable and the different school characteristic variables.

Model	Variables	*Sw*	*En*	*Ma*	*ComGr*	PAREDU covariance	χ^2 df	RMSEA
D1	INDEPSCH[1]	.05[ns]	.09	.08	-.04[ns]	.44	4792.85	.024
	PAREDU	.70	.76	.72	-.32		82	
D2	SIZE[2]						5123.73	.023
	300-499	-.06[ns]	-.06[ns]	-.05[ns]	-.11[ns]	-.07	94	
	500-1000	-.04[ns]	-.03[ns]	-.05[ns]	-.13[ns]	.10		
	>1000	.04[ns]	.03[ns]	.01[ns]	.01[ns]	.11[ns]		
	PAREDU	.71	.79	.75	-.32			
D3	LOCATION[3]						4944.30	.024
	Rural	-.11	-.07	.11	.02[ns]	-.12	88	
	Suburban	.12	.06	.16	.19	.28		
	PAREDU	.76	.82	.80	-.26			
D4	TEAAGE	-.01[ns]	-.05[ns]	.00[ns]	.08	-.11	4970.79	.025
	PAREDU	.71	.80	.74	-.33		82	
D5	TEAEXP	-.02[ns]	-.05[ns]	.02[ns]	.04[ns]	-.15	4948.01	.024
	PAREDU	.71	.79	.75	-.33		82	
D6	TEACER	.01[ns]	-.02[ns]	.05[ns]	-.03[ns]	.07[ns]	4846.26	.024
	PAREDU	.72	.80	.74	-.31		82	
D7	TEAWOM	-.01[ns]	-.04[ns]	-.02[ns]	-.07[ns]	.23	4853.90	.024
	PAREDU	.72	.79	.75	-.33		82	
D8	TEADENS	.04[ns]	.05[ns]	.05[ns]	-.03[ns]	.17	4868.46	.024
	PAREDU	.71	.79	.73	-.34		82	

Note. [1] Municipal schools are the reference. [2] Size <300 is the reference. [3] Urban schools are the reference. All values are significant on the .001 level except for the marked values (ns) which are non-significant.

The first model (model D1) showed a good fit (χ^2 (82, 99070)= 4792.85; RMSEA=.024) and showed that there is a strong, positive covariance between PAREDU and INDEPSCH (.44) which shows that, in comparison to municipal schools, in independent schools students have well-educated parents to a greater extent. The strength of the regression coefficients between INDEPSCH and the subject-specific dimensions also decreased substantially, and become insignificant for the *Sw* factor. These results show that parental education confounds the relation between INDEPSCH and grades (see Figure 2).

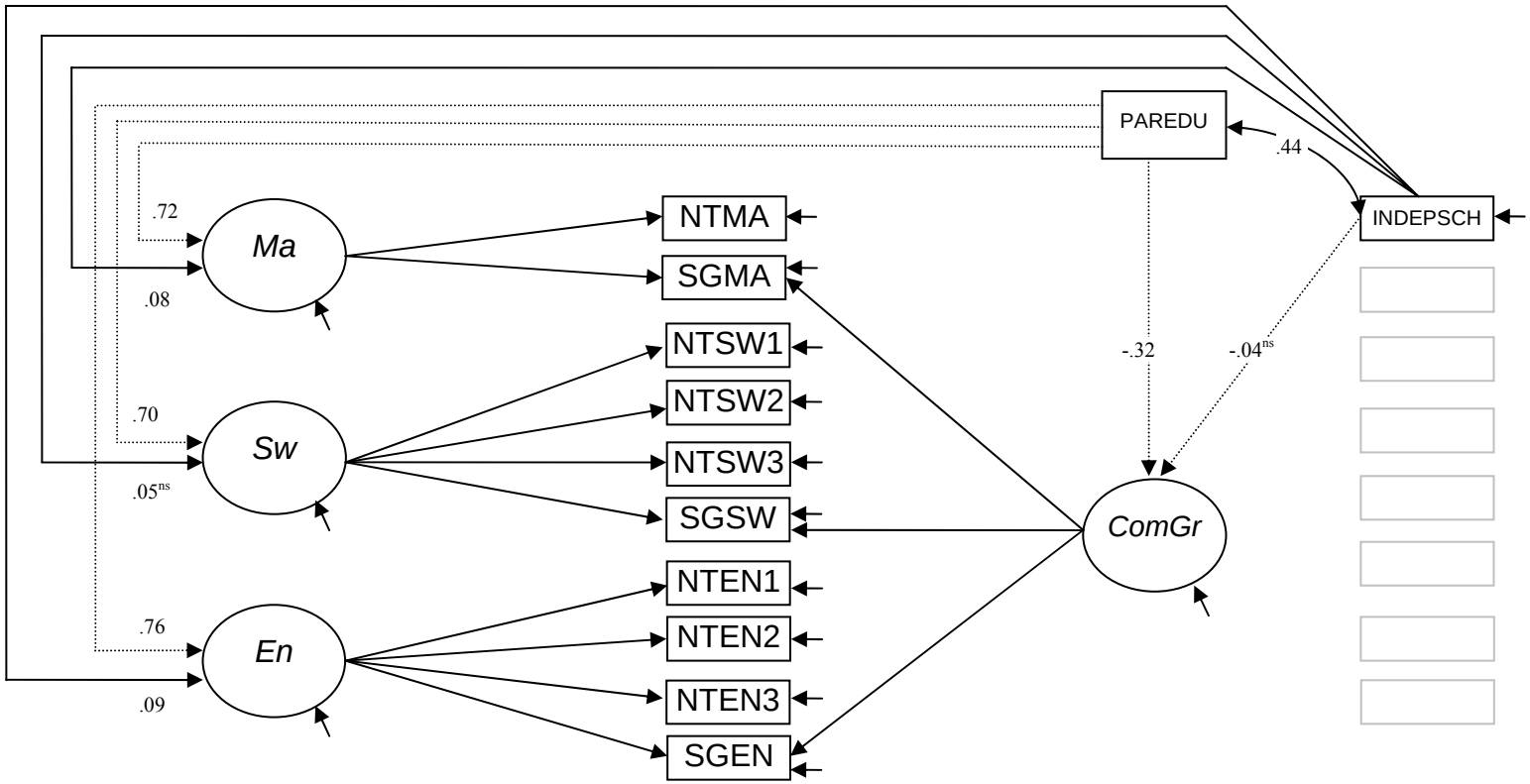

Figure 2. The structural two-level models (D1-D8) with relations between one student characteristic factor and the subject-specific and common grade factors, and with covariances between the residuals for *Sw*, *En* and *Ma*. Parental education as a dummy variable with covariances between the different school characteristics and parental education, at the school level. The small arrows are residuals. The estimates given are an example of the relations in model D1, see Table 3.

Model D2 also showed a good fit (χ^2 (94, 99070)= 5123.73; RMSEA=.023) and there was a significant covariance between PAREDU and SIZE for the second largest and largest schools (-.07, .10, respectively), in comparison to the reference category small schools. All the regression coefficients between the subject-specific and common grade factors (*Sw*, *En*, *Ma, ComGr*) and SIZE became non-significant when parental education was controlled for.

The model with covariance between LOCATION and PAREDU (model D3) also showed a good fit (χ^2 (88, 99070) = 4944.30; RMSEA=.024) and that there was a significantly negative covariance between PAREDU and schools located in rural areas (-.12) whereas a significantly positive covariance was found for PAREDU and schools located in suburban areas (.28), in relation to the reference category urban schools. These results show that in rural schools, students have less well-educated parents in comparison to suburban and urban schools. The model also showed that when PAREDU was controlled for, the relations between the subject-specific factor *Sw* and *En* and LOCATION became significantly negative for schools located in rural areas (-.11, -.07, respectively) and significantly positive for schools located in suburban areas (.12, .06, respectively). For the *Ma* factor, schools located in rural as well as suburban areas had significantly positive values (.11, .16, respectively), in comparison to the reference category urban schools. The association between schools located in suburban areas and *ComGr* decreased somewhat but still remained strong when PAREDU was controlled for (.19).

The results for the models with TEAAGE (model D4) and TEAEXP (model D5) showed that the model fit was good for both models (χ^2 (82, 99070) = 4970.79; RMSEA=.025) and (χ^2 (82, 99070)= 4948.01; RMSEA=.024), respectively. These models had similar results, namely that there were negative covariances between TEAAGE, TEAEXP and PAREDU, which indicates

that schools with older and more experienced teachers have students with less well-educated parents. The model with TEACER (model D6) showed no significant relations. Similar results were also found in the final two models, where there were significant and positive covariances between TEAWOM (model D7) and TEADENS (model D8) and PAREDU. These results indicate that schools with a large proportion of female teachers and schools with a high teacher density have students with well-educated parents to a greater extent.

To sum up, when parental education was controlled for the results showed that the strong influence of INDEPSCH on the subject-specific factors decreased substantially and that all the associations between SIZE and the subject-specific and common grade factors became non-significant. When parental education was controlled for, the non-significant association between rural schools and the subject-specific factors became significant. All the associations between the teacher characteristic variables and the subject-specific factors also became non-significant when parental education was controlled for, except for the association between TEAAGE and the common grade factor which instead became significantly positive.

Discussion and conclusions

The purpose of this study was to investigate systematic school differences in the subject-specific dimensions and the common grade dimension in grades. Different school characteristics were related to a previously estimated two-level four-factor model (Klapp Lekholm & Cliffordson, 2008). Since the previous study had found that parental education influenced the subject-specific dimensions and the common grade dimension rather considerably and, in reversed direction, on the school level, parental education was controlled for in order to investigate the confounding relations between parental education and different school characteristics in grades.

The main finding of this study is that whilst some school characteristics have effects on grades, when parental education was controlled for, most of these effects are reduced and become non-significant. Thus, the results show that there are confounding associations between parental education and the different school characteristics and grades, and that it is not, for example, the provider of schools, school size or school location per se that influence the subject-specific and common grade dimensions in grades. Parental education influences the relation between independent schools and grades quite substantially, as well as the similar relations between grades and school size, location and the different teacher characteristics.

Municipal and independent schools

The finding that the 'independent school effect' is primarily a 'parental education effect' supports the contention that research concerning this issue demands appropriate methods that can both control for student background and take account of the multilevel structures in the data (Gustafsson & Myrberg, 2002; McEwan, 2000; Somers et al., 2001). The finding that pupils at independent schools gain higher grades is, to a large extent, a result of the peer characteristics of the schools; independent schools have a higher proportion of students with well-educated parents. Thus, the result of this study contradicts previous research which claims that independent schools award higher grades per se (Wikström, 2005).

School size and school location

The findings in previous studies (Darling-Hammond et al., 2006; Wikström, 2005) that small schools grade more highly do not accord well with the result in this study, which instead suggests that the effect of school characteristics such as school size and location may be due to the confounding influences of parental education. The initial analyses showed that, in comparison to smaller schools, large schools award higher grades, but, when parental

education was controlled for, the effect of large schools on the two grades dimensions became non-significant. This supports some of the results of previous research which suggest that it is not school size per se that influences achievement, but demographic factors (Darling-Hammond et al., 2006; Lee & Smith, 1997; Ready & Lee, 2006). It thus seems reasonable to believe that the relation between school size and grades is a function of student background and demographic variables. The conclusion of Darling-Hammond et al. (2006) that small schools produce better results is a finding that does not receive confirmation in the present study. However, it should be noted that a large school in a Swedish context might be considered a small school in another country.

Schools located in suburban areas grade higher in comparison to schools located in rural and urban areas. When parental education is controlled for, suburban schools still award higher grades but the influence of the common grade dimension in grades decreases. This result may be due to an association between school size, location of schools and parental education, where large schools are primarily located in suburban and urban areas where parents are, to a greater extent well-educated, and that rural schools may be small schools by default where parents are less well-educated. This result shows confounding associations between rurality, school size and parental education. The result also indicates that the peer characteristics may differ quite substantially among schools located in suburban areas, which may be due to segregation effects. The current voucher system for compulsory education in Sweden, offers families the possibility to choose the school for their children, which implies the development of a segregated school system where the underlying social structure affects the educational system to function as a tracking system. However, in Sweden, formal tracks do not exist in compulsory school. Stanley et al. (2008) and Williams (2005) claim that it is the

characteristics of the neighbourhood and the underlying structures in society that influence achievement and grades.

Teacher characteristics

Without controlling for parental education, relations between the different teaching characteristics and the subject-specific dimensions in grades can be found. The initial analyses showed that schools with younger and less experienced teachers award higher grades and that teacher certification has only a significant relation to the mathematics dimension. Schools with a higher proportion of female teachers and higher teacher density register higher achievement on the subject-specific dimensions. Teacher characteristics have no significant relation to the common grade dimension. However, when controlling for parental education the relation also became non-significant for the subject-specific dimensions, which may be due to confounding effects such as the compensatory allocation of resources and selection effects among teachers. The non-significant result of, for example, teacher certifications on grades, may be due to an overly broad instrument or variables being used which do not manage to separate specific teacher characteristics of importance for student achievement and grades.

Whilst Greenwald et al. (1996) claim that teacher effectiveness has strong associations with teacher characteristics, Hanushek (1989, 1997) argues that no measurable teacher characteristics that influence teacher effectiveness can be found. The results of the current study show no clear associations between teachers' certifications and student grades. However, different factors may obstruct the associations of different school characteristics and grades. The estimates of relations between teachers' subject knowledge and pedagogical competencies and certifications may be affected by reliability and validity problems of the

instrument measuring teacher characteristics. One explanation is that the instrument does not reflect important and detailed aspects of teacher competence, such as for example if an uncertified teacher is uncertified because of a few missing credits from her/his teacher education, but is well qualified in a subject area of importance for instruction. The importance of the formal certification in teacher education is probably confounded on the one hand by factors not controlled for and, on the other, due to instruments that lack the capacity to measure relevant characteristics.

Conclusions

Parental education is a factor which exerts a strong influence on school achievement and grades, and which is correlated with different school characteristics. It therefore affects estimates of relations between school characteristics and grades in different confounding ways. The characteristics of the peer-group are probably of major interest in order to understand why parental education has a strong influence on grades. Schools with many students with parents who have low levels of educational attainment benefit in the common grade dimension, which may be interpreted as reflecting a compensatory grading practice, in that teachers award grades which are higher than the students' performance on the national tests.

The result that the different teacher characteristics do not influence the common grade dimension suggests that this dimension is universal and robust against the influence of different teacher characteristics such as age, experience and certifications. However, the fact that there is a common grade dimension may be an expression of the moral dilemmas teacher face when assessing and grading their students. Deutsh (1979) and Pilcher (1994) argue that teachers are concerned with the distributive justice of grades, which is not always in line with

the curriculum. The dilemma surrounding the intersection between the ethical and moral issues and the pressure for objective grading that teachers face (Wiliam, 1996) may lead to grading structures that are opaque and difficult to recognize, thus making it hard to assess the validity of grades. However, the prognostic validity of the current grading system in Sweden has been shown to be as good as, or indeed even better than, for the previous system of norm-referenced grades (Cliffordson, 2008). Interestingly enough, the norm-referenced grading system was constructed primarily to function as a selection instrument to the next level in the educational system, whereas the current criterion-referenced grading system was primarily constructed in order to give information about student attainment.

Grades reflect several dimensions of student achievement and the non-achievement dimension in grades may be of major importance for students in order to succeed in the educational system and in society at large. However, the curriculum and policy documents do not recognize this grading practice, thus obscuring the transparency concerning the types of knowledge and characteristics that grades actually measure. The issue of validity in grades would thus appear to be of major interest in order to understand the actual grading procedure and to evaluate the function of the grading system.

Methodological issues

The results stress the need for appropriate methods when conducting studies involving data from different levels, for example student and school characteristics. In order to handle the multilevel structure of school data, multilevel analyses must be conducted. Another issue concerns study design, where the drawing of causal inferences from cross-sectional studies can be questioned. Further, it has been stressed by several researchers that the control of student peer groups that are of major importance, has been neglected in many studies (Somers

et al., 2001). Some researchers have pointed at the difficulties of making sound interpretations of results when inappropriate methods of analyses have been used (Gustafsson & Myrberg, 2002; Hanushek, 1989; Lee & Smith, 1997; McEwan, 2000; Somers et al., 2001).

Limitations

One limitation of the current study is that the investigated school characteristics are somewhat limited, and there may be several other school characteristics which could also influence grades. The instrument measuring teacher competencies may also be imprecise in that it may not capture important and relevant aspects of teacher competencies and education. Another limitation concerns the use of national tests as an instrument of measuring achievement since the tests are assessed and graded by the teachers themselves. It seems reasonable to believe therefore that the tests also measure a certain amount of non-achievement. However, the procedure around the assessment and grading of the tests has been of major interest and extended procedures around the tests have been developed during recent years which secure the validity and reliability of the tests.

Acknowledgement

Swedish Research Council has financially supported the research reported in this paper. The work is a part of the GRAM (Grade and Grade Assignment: Functions and Effects) project.

References

Bergström, F., & Sandström, M. (2001). *Konkurrens bildar skola – en ESO-rapport om friskolornas betydelse för de kommunala skolorna*. [Competition and the Quality of Municipal Schools] Ds 2001:12. Stockholm, Sweden: Ministry of Finance.

Cliffordson, C. (2008). Differential prediction of study success across academic programs in

the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment, 13*(1), 56-75.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartlant, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U. S. Government Printing Office.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state of policy evidence.* Washington, DC: U. S: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, *51*(3), 166-173.

Darling-Hammond, L., Ross, P., & Milliken, M. (2006). High school size, organization, and content: What matters for student success? Brookings Papers on Educational Policy: 2006/2007.

Deutsch, M. (1979). Education and distributive justice. *American Psychologists*, *34*(5), 391-401.

Figlio, D. N., & Stone, J. A. (1999). School choice and student performance: Are private schools really better? Working paper 1141-97, California: Institute for Research on Poverty, University of Oregon.

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size: Research and Policy.* Beverly Hills, CA: SAGE Publications.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361-396.

Gustafsson, J.-E., & Myrberg, E. (2002). Ekonomiska resursers betydelse för pedagogiska resultat [The importance of economical resources on pedagogical results]. Stockholm: Liber.

Gustafsson, J.-E., & Stahl, P.-A. (2005). *STREAMS 3.0 User's Guide*. Mölndal, Sweden: Multivariateware.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance, *Educational Researcher*, *18*(4), 45-65.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An Update. *Educational Evaluation and Policy Analysis*, *19*(2), 141-164.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143-163.

Jöreskog, K. G. (1993). Testing Structural Equation Models. In K. A. Bollen, & J. Scott Long (Eds.), *Testing Structural Equation Models* (pp. 294-316) California: Sage Publications.

Klapp Lekholm, A., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation*, *14* (2), 181-199.

Klapp Lekholm, A., & Cliffordson, C. (in press). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best, and for whom? *Educational Evaluation and Policy Analysis, 19*(3), 205-227.

Lindahl, M. (2005). *Studies of causal effects in empirical labor economics*. Dissertation, Stockholm: SOFI.

McEwan, J. P. (2000). Comparing the effectiveness of public and private schools: A review of evidence and interpretation. Occasional paper presented at teachers college, Columbia University, New York, National Centre for the study of privatization in education, June 2000.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modelling with data that are not missing completely at random. *Psychometrica, 52*(3), 431-462.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus User's Guide* (3rd edition). Los Angeles, CA: Muthén & Muthén.

National Agency for Education. (2004). *Likvärdig bedömning och betygssättning. Skolverkets allmänna råd 2004* [Equal assessment and grading. General advice issued by the National Agency for Education]. Stockholm: National Agency for Education.

National Agency for Education. (2007a). Provbetyg – slutbetyg – likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolans årskurs 9, 1998-2006. [Test grade – leaving certificate – equal assessment? A statistical analysis of the relation between national tests and final grades in compulsory school, year nine 1998-2006]. Stockholm: National Agency for Education.

National Agency for Education. (2007b). National subject tests in compulsory school. Retrieved May 2, 2007 from http://www.skolverket.se/sb/d/276.

National Agency for Education (2008). Compulsory School. Retrieved June 10, 2008 from http://www.skolverket.se/sb/d/354/a/959#paragraphAnchor3

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, *2*(1), 69-88.

Ready, D. D., & Lee, V. E. (2006). Optimal Context size in elementary schools: Disentangling the effects of class size and school size. *Brookings Papers on Education Policy*, 2006/2007, 99-135.

Roscingo, V. J., & Crowley, M. L. (2001). Rurality, institutional disadvantage, and achievement/attainment. *Rural Sociology*, *66*(20), 268–292.

Schafer, L. J., & Graham, W. J. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Somers, M.-A., McEwan, P. J., & Willms, D. J. (2001). How effective are private schools in Latin America? Occasional paper, New York: National Center for the Study of Privatization in Education, Columbia University.

Stanley, R. L., Comello, M. L. G., Edwards, W. R., & Marquart, S. B. (2008). School adjustment in rural and urban communities: Do students from "Timbuktu" differ from their "City slicker" peers? *Youth Adolescence*, 37, 225-238.

Swedish National Rural Development Agency. Retrieved March 4, 2008 from http://www.glesbygdsverket.se/site/default.aspx.

Wikström, C. (2005). *Criterion-Referenced Measurement for Educational Evaluation and Selection. Dissertation No 1*. Umeå: Umeå University, Department of Educational Measurement.

Wiliam, D. (1996). National curriculum assessments and programmes of study: Validity and impact. *British Educational Research Journal, 22*(1), 129-142.

Williams, S. L. (2001). City kids and country cousins: Rural and urban youths, deviance, and Labor market ties. In R. T. Michael (Ed.), *Social awakening* (pp. 379–414). New York: Russell Sage Foundation.

Åberg-Bengtsson, L., & Erickson, G. (2006). Dimensions of national test performance: A two-level approach. *Educational Research and Evaluation*, *12*(5), 469-488.