Educational Research and Evaluation
Vol. 00, No. 0, Month 2008, 1–22

Routledge
Taylor & Francis Group

Effects of student characteristics on grades in compulsory school

Alli Klapp Lekholm[a]* and Christina Cliffordson[a,b]

[a]*Department of Education, University of Gothenburg, Sweden;* [b]*University West, Trollhättan, Sweden*

(*Received 31 January 2008; final version received 10 June 2008*)

The purpose of the study was to investigate how different student characteristics such as gender influence grades. In order to answer these questions, multivariate techniques were used. The data derive from The Gothenburg Educational Longitudinal Database (GOLD), and the subjects were 99,070 ninth-grade students born in 1987. The analyses were performed on subject grades and scores on national tests in Swedish, English, and mathematics and on student questionnaire data. The results showed that the greatest part of the variance in grades was due to student achievement in the different subject areas but that grades are additionally influenced by different noncognitive student characteristics, such as motivation and parental engagement. One of the most interesting results concerns the relation between student motivation, gender, and grades.

Keywords: compulsory school; gender; grades; grade assignment; motivation; national tests; student characteristics

Introduction

A substantial body of research shows that student characteristics and teachers' different grading practices exert an influence on the assignment of grades (Alexander, 1935; Brookhart, 1991, 1993, 1994; Cliffordson, 2004; Klapp Lekholm & Cliffordson, 2008; Pilcher, 1994; Wikström, 2005). Research on student characteristics of importance for achievement has indicated that student self-efficacy and motivation are two major characteristics influencing achievement and grades (Brookhart & Devoge, 1999). It has also been suggested that grades reflect student social behaviour as well as academic knowledge, and that cooperation may be critical for school success and grades (Lane, Givner, & Pierson, 2004; Wentzel, 1989).

The relation between cognitive ability and achievement is strong (Gustafsson & Balke, 1993) although different noncognitive factors, such as family values and processes, out-of-school experiences, and patterns of attitudes and motivation, all seem to be of importance for understanding achievement differences (Wang & Lin, 2005). In order to gain an understanding of the causes of differences in grades, cognitive as well as noncognitive factors are thus of vital importance (Gipps & Murphy, 1994; Murphy & Whitelegg, 2006). In school, students are expected to manage their academic, social, and behavioural performances, as well as to adapt their performance in relation to the expectations of teachers, parents, and society at large.

*Corresponding author. Email: alli.klapp.lekholm@ped.gu.se

ISSN 1380-3611 print/ISSN 1744-4187 online

DOI: 10.1080/13803610802470425
http://www.informaworld.com

In Sweden, school grades are assigned by teachers, primarily based on classroom assessments, where they judge the performances of their students on a daily basis, and by the use of teacher-made tests. In order to help the teachers calibrate their grading against the criteria in the curriculum, national tests are used in three core subjects: Swedish, English, and mathematics. Although the national tests are centrally produced, they are distributed, assessed, and marked by the teachers themselves. Teachers have access to a national test bank containing a large amount of national tests and examples of student performance levels. The educational system relies on teachers' assessments of student performances to be fair. Since grades are used for selection purposes, for example, for entrance to the next level in the educational system, it is an important assumption that grades are comparable, reliable, and valid measures of student achievement.

Klapp Lekholm and Cliffordson (2008) identified two dimensions in grades within and between schools: a subject-specific dimension relating to student academic achievement as measured by the results on national tests and grades and a common grade dimension related only to grades, which cuts across grades in different subjects and is interpreted as relating to noncognitive factors. The common grade dimension suggests that grades are influenced by factors other than academic achievement. Whilst the influence of student gender was found to be significant in the common grade dimensions within schools, parental education level was not found to be associated with the common grade dimension within schools.

Gender may influence the common grade dimension in different ways. One possibility is that gender has a direct influence on the common grade dimension in grades, while another possibility is that gender operates through different student characteristics and thus has an indirect effect on grades. This implies that noncognitive factors, such as student interest and motivation, parental engagement in the school work, and self-efficacy, might function so as to mediate gender differences.

In this study, the main aim is to investigate which student characteristics influence the subject-specific and common grade dimension in grades. A primary focus is placed on the common grade dimension and the incorporation of measures of individual characteristics such as student interest, adjustment, self-efficacy, cooperation, and the engagement of parents, in order to more closely clarify and define the meaning of the common grade dimension. An additional aim is to explain gender differences in grades in terms of the mediating effects of student characteristics on grades.

Previous research

Certain student characteristics, such as motivation and effort, may be seen as sources of relevant variance in grades, in that these characteristics influence learning, while other characteristics which affect grades without affecting learning could be regarded as sources of irrelevant variance in grades (Messick, 1994; Wiliam, 1996). The association between assessment and grading practices and validity is of major importance. The distributive justice in grades (Deutsch, 1979; Pilcher, 1994) seems to be in focus for teachers where the relevance of a grade may depend on its consequences and how it is used, which puts demands on the consequential validity aspects.

In Sweden, grades from upper secondary education have been shown to be a good predictive instrument for student achievement in higher education, which may be due not only to the reflection of achievement differences in grades but also of noncognitive factors (Cliffordson, 2008; Gustafsson & Balke, 1993). While the constructs or characteristics that are measured by grades might be irrelevant according to the grading criteria and the

grading system, they may, however, be relevant for how grades function in the educational system and in society at large. Several studies have found that the awarding of grades is influenced by a tension between teachers' internal beliefs and values and external influences and pressures on them (McMillan; 2003; Wiliam, 1996). In sum, it may be that construct-irrelevant variance in grades may be relevant for predicting student achievement in the educational system (Klapp Lekholm & Cliffordson, 2008).

Student characteristics of importance for grades

Several studies have investigated teachers' expectations of student behaviour and the types of student characteristics that are of importance for grades. According to Wentzel (1991), student characteristics such as adaptive behaviours and behavioural and interpersonal forms of competence are strong predictors of academic success and influence grades. Wentzel (1989) also found that adolescents' grade point average was positively related to behavioural and social competencies, such as effort and motivation, and whether students were responsible and endeavoured to understand things. A study conducted by McMillan (2003) showed that teachers used student effort as a key criterion for students on the board for higher grades. Teachers' assessment and grading decisions seem quite strongly based on the beliefs of what best would result in student motivation and engagement.

Wentzel (1991) investigated relations between grades and three aspects of social competence among 12-year-olds and found that socially responsible behaviour, such as cooperation, respect for others, and positive forms of group participation, was most relevant to high grades. When controlling for background variables such as IQ, gender, ethnicity, and family structure, the socially responsible behaviour mediated almost the entire relationship between students' grades and social competence. Wentzel (1991) concluded that socially responsible behaviour in the classroom contributes directly to learning and academic achievement. She also concluded that boys and girls pursue goals of both academic and social competence with equal frequencies and she argued that "whereas some sex differences in classroom goal pursuit might exist, it is not reasonable to assume that male and female classroom competence (at least at the upper end of the continuum) is related to gender-specific goals" (p. 140). However, according to other research, students have individual schemas, some of which are closely related to the goals of classroom learning, whilst others are not (Arnot, David, & Weiner, 1996; Murphy, 2000).

Research suggests that student interest is characterized by varying amounts of affect, knowledge, and value and thereby includes both cognitive and noncognitive aspects (Hidi & Renninger, 2006). It has also been argued by several researchers that individual and situational interests are two types of interest that are motivational since they concern perception and actions that a person considers possible (Bergin, 1999; Brookhart & Durkin, 2003). According to Ainley, Hidi, and Berndorff (2002) individuals have specific individual interests, some of which may be in line with the goals of learning in the classroom, whilst others are not. It is also argued that students may have domain-specific interests as well as a more general individual interest in learning (Pintrich, 2002). Students who develop a general interest in learning often possess both specific and general goals and approach new phenomena with the goal of understanding them. It has also been recognized that interest and motivation are closely related (Brookhart & Durkin, 2003).

Research has shown that self-efficacy is related to student effort, persistence, and performance (Brookhart, Walsh, & Zientarski, 2006) where students make judgements about their own self-efficacy by judging the difficulty of a task, the amount of effort that is required to solve it, and the amount of assistance they need. Research on the influence of

anxiety or low self-efficacy on student achievement has consistently demonstrated that student anxiety causes poor achievement and is related to students' fears of negative evaluations and to student defensiveness (Giota, 2006; Hembree, 1988).

In summary, the research reviewed here shows that cognitive as well as noncognitive factors, such as student motivation, interest, self-efficacy, and cooperation, influence grades.

Gendered student achievement

In most industrialized countries, girls perform better in school and obtain better grades in comparison to boys (National Agency for Education, 2003). By the end of compulsory school, girls in Sweden have better grades in all subjects with the exception of physical education. Differences in grades are smallest in mathematics, physics, and technology and greatest in Swedish, art, home economics, and religion (National Agency for Education, 2005). Gender differences in teacher-awarded grades are much greater in comparison to results on the national tests. The gender differences in the national tests themselves are most apparent in the test for Swedish, moderate in the English test, but negligible in the national test for mathematics.

Andersson (1998) fitted a hierarchical model in order to analyse the dimensionality of grades in compulsory school. The result indicated that a general school-achievement factor influences every single grade, and it was interpreted to be a strong factor. However, the interpretation of this school-achievement factor was not clear, but it was argued that it could be composed of different student characteristics such as motivation and school adjustment. Girls had a higher mean on this general school-achievement factor. Gustafsson and Balke (1993) also identified a strong general school-achievement factor in grades and found this factor to correlate around .60 with general cognitive ability measured 3 years earlier. Thus, while the general school-achievement factor to a certain degree overlaps with general cognitive ability, a large proportion of variance in achievement, as reflected in grades, is independent of cognitive ability.

Research has also shown that girls come to school better equipped for school activities such as writing and drawing in comparison to boys (Arnot et al., 1996; Murphy, 2000). It also seems as if girls' and boys' different interests make girls better prepared for school and that there is a close correspondence between school achievement and the types of pre-school activities that girls prefer (Murphy, 2000). Moreover, these different learning approaches may affect and relate to other student characteristics, such as interest, motivation, parental engagement, and perception of self-competence and social skills, and thus have an influence on grades.

Gendered student non-achievement, experiences, and expectations

According to Murphy (2000), girls and boys develop different behaviours due to the experiences and expectations they meet, which in turn leads to different perceptions of themselves, of others, and of the social environment. These different expectations will often lead them towards different experiences – both within and out of school – such as perceptions of male and female domains and of relative competencies. Teachers have been found to expect and judge the performances of girls and boys in stereotyped ways, where girls are expected to perform better in language domains while boys are believed to do better in quantitative domains. However, these presumptions may be more related to students' behavioural and social differences than to actual achievement (Gipps & Murphy,

1994; Rosén, 1998). Out-of-school experiences give students different opportunities to perform in school; for example, boys have been shown to outperform girls on achievement in tasks related to their interests, such as using instruments for weighing and measuring (Murphy, 2000). Additionally, girls' and boys' different experiences and interests may influence assessment outcomes and show real gender differences due, for example, to their different views about "relevance", and which may not indicate any differences in actual competence.

Student interest has been shown to be an important cognitive and emotional resource for school outcomes (Hidi, Renninger, & Krapp, 2004). Since research has found that the influence of gender on grades is significant at the individual level, it seems reasonable to investigate gender differences and the ways in which gender affects different student characteristics and how these different characteristics mediate gender differences in grades.

Purposes

The review of the research literature indicates that there is reason to believe that grades may be systematically affected by student characteristics of which some influence grades through learning, some influence grades in spite of the fact that they do not influence learning, and some influence grades in both ways. The main aim of the study is to investigate the direct and indirect relations between student characteristics and grades. Another aim is to investigate the influence of gender in grades in terms of mediating factors.

By using student questionnaires, grades, results on national tests, and data on gender, it is possible to use multivariate techniques to investigate how different student characteristics influence grades. Specifically, the following questions will be addressed:

- Which student characteristics are related to grades, and how do different student characteristics influence the subject-specific and common grade dimensions, respectively?
- How does gender influence grades, and, specifically, how do different student characteristics mediate the effect of gender on grades?

Method

Participants

Data used in this study come from The Gothenburg Educational Longitudinal Database (GOLD), which contains register data compiled by Statistics Sweden for all individuals born between 1972 and 1987 and where a large amount of information, such as gender, parents' education, grades from compulsory and post-compulsory education, and results on national tests, is available. The subjects in the study were 99,070 students born in 1987 who had left compulsory school in 2003. The data used are subject grades, national test results from the final grade in compulsory school, and gender. The student questionnaire data derive from the Evaluation Through Follow-up project (ETF, see Härnqvist, 2000) where a 10% nationally representative sample has been collected. However, two reductions of the group of participants have been made. First, individuals for whom both subject grades and national test results are lacking have been excluded. Secondly, schools with 14 students or fewer have been excluded, since extremely small schools are often schools for students with special needs. In sum, 1,782 individuals have been excluded from the analyses.

Instruments

Three types of measures were used, namely subject grades, national test scores in Swedish, English, and mathematics, and a student questionnaire. The student questionnaire consisted of items grouped into different areas related to student characteristics.

The Swedish grading system

Several major school reforms took place in Sweden in the beginning of the 1990s, which led to a highly decentralized educational and grading system. The present grading system is based on criterion-referenced grades and an underlying assumption that teachers will interpret these criteria in a similar way. The grading is primarily based on classroom assessment. The purpose of the grading system is to provide information about the individual student's acquisition of required standards, to evaluate the system, and to be used as an instrument for selection to the next educational level in the school system. Individual teachers judge the performance of their students and grade them. National tests are used in order to help the teachers calibrate their grading. In Sweden, there are no tracks in compulsory school. The different subject domains are defined by the curriculum and by specific criteria. In the Swedish national curriculum, the grading criteria for Swedish and English often reach beyond the respective subject domains. Some criteria emphasize, for example, active communication, the development of thinking and argumentation, and the ability to advance opinions, while other criteria deal more specifically with subject content, such as grammar and vocabulary. However, in mathematics, the criteria primarily emphasize subject content, for example, that students should demonstrate knowledge in areas such as geometry, diagrams, and equations (National Agency for Education, 2007c).

The grading scale encompasses four levels: not pass (IG), pass (G), pass with distinction (VG), and pass with special distinction (MVG). Since the grades are used for selection purposes, they are converted into four numbers 0, 10, 15, 20: not pass (IG) = 0, pass (G) = 10, pass with distinction (VG) = 15, and pass with special distinction (MVG) = 20. No intermediate numbers are used. These levels reflect student performance in relation to the objectives for each subject. In this study, grades from the end of Grade 9 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 have several purposes, for example to function as instruments enabling teachers to calibrate their grading and to be used in order to drive curriculum implementation (Åberg-Bengtsson & Erickson, 2006). 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 criteria are tested, which implies that the different subject domains are not fully covered by the tests. In Grade 9, national tests are used in three core subjects: Swedish, English, and mathematics. In Swedish, there are three subtests, the first testing reading comprehension, the second an oral test conducted in pairs of students, and the final subtest 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, 2007a). There are no external referees involved in the assessment and grading procedures of the tests. However, in order to secure the validity and reliability of the tests, teachers are strongly recommended to cooperate in the assessment and grading process of the tests. Moreover, the National Agency for Education has developed a test bank where teachers have access to a large amount of national tests and examples of different levels of student achievement. The National Agency for Education (2007b) has conducted several studies and argues that the national tests are useful instruments in order to investigate the equivalence in grades.

The scale for the national tests corresponds with the scale for the 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 only scale points used, no intermediate scores being used. In this paper, the following abbreviations are used for the national tests: NTSW1, NTSW2, and NTSW3 for the test scores in Swedish, for English NTEN1, NTEN2, and NTEN3 are used, whilst for mathematics only one summarized test score is available, NTMA.

Questionnaire data

The questionnaire data were collected during spring 2003, when the students were finishing their final year of compulsory school (Grade 9). Out of a large number of items, 43 items representing certain student characteristics have been used as indicators to create latent variables or factors (Table 1). In all, eight factors have been created. Three self-perceptions of competence factors – *SpSw*, *SpEn*, and *SpMa* – were created, which were hypothesized to reflect students' self-perceptions of competence, interest, and self-esteem in the three subjects Swedish, English, and mathematics. One factor designed to reflect parental engagement in the school work of their child (*Parent*) was created and was related to four items concerning whether parents and children talk about school and school work at home, how the child is progressing, and whether the parents are participating in school activities. Another factor was hypothesized to reflect student self-efficacy and to function as a measure of coping in school (*Coping*), which was generated by four indicators focusing on student self-perceived experiences of school, for example, whether the student received help when needed. Another factor was hypothesized to reflect students' general interests (*Interest*) in learning more in the different subjects and was measured by three indicators about interest in the respective subjects and one indicator about considering future university study hypothesized to reflect student general interest or motivation for education and further studies. Yet another factor was hypothesized to reflect student adjustment and happiness in school (*Adjust*) and was measured by five indicators about how content the student was with peers, teachers, and school work. The final factor was hypothesized to reflect student cooperation (*Cooperate*) with peers and teachers in the classroom and was measured by three indicators.

Methods of analysis

In order to investigate the influence of student characteristics and gender in the subject-specific and the common grade dimensions, confirmatory factor analysis (CFA) and structural equation modelling (SEM) were used. In a previous study (Klapp Lekholm & Cliffordson, 2008), multilevel CFA and SEM were used in order to account for school

Table 1. Student characteristics measured by questionnaire data.

Items	Answers
SpSw – Self-perceptions of competence in Swedish	
1. How well do you think you are able to accomplish the following in Swedish? a. Read and understand a text b. Read aloud for the whole class c. Read the subtitles to a film d. Write a story e. Participate in a discussion f. Give an oral presentation	Very bad – very good (5-point scale)
2. How good do you think you are in Swedish?	Very bad – very good (5-point scale)
9a. How interested are you in learning more about Swedish?*)	Not at all – very interested (4-point scale)
SpEn – Self-perceptions of competence in English	
3. How well do you think you are able to accomplish the following in English? a. Understand when someone speaks b. Understand an English film without subtitles c. Read and understand a text d. Talk to anybody e. Write a letter or a story f. Look up factual information in an English encyclopaedia g. Give an oral presentation	Very bad – very good (5-point scale)
4. How good do you think you are in English?	Very bad – very good (5-point scale)
9b. How interested are you in learning more about English? *)	Not at all – very interested (4-point scale)
SpMa – Self-perceptions of competence in mathematics	
5. How well do you think you are able to accomplish the following in mathematics? a. Mental arithmetic/rough calculation b. Calculate addition and division c. Calculate percentages d. Calculate area and circumference e. Solve equations f. Solve mathematic problems g. Explain mathematics problems	Very bad – very good (5-point scale)
6. How good do you think you are in mathematics?	Very bad – very good (5-point scale)
9c. How interested are you in learning more about mathematics? *)	Not at all – very interested (4-point scale)
Parent – Support from parents	
7. How often do: a. you tell your parents about how it has been or how you have felt in school? b. you tell your parents about what you have been working with in school? c. you show your parents your test results or other things you have done in school? d. your parents participate in school?	Never – very often (5-point scale)

(*continued*)

Table 1. (*Continued*).

Items	Answers
Coping – Coping	
8. Do you: a. find it difficult to keep up in lessons? b. give up if you get a difficult task to do in school? c. need more help than what you get from your teachers? d. often find it difficult to concentrate in lessons?	Always – never (5-point scale)
Interest – Interest	
9.*) How interested are you in learning more: a. about Swedish? b. about English? c. about mathematics?	Not at all – very interested (4-point scale)
10. Are you considering studying at university?	No, not interested – not decided yet – yes
Adjust – Adjustment	
11. How content are you: a. in your current class? b. in your current school? c. with other pupils? d. with the teachers? e. with your school work?	Very bad – very good (5-point scale)
Cooperate – Cooperation	
12. How often: a. do the teacher and students discuss together in your class? b. do you work in groups in your class? c. are students involved in planning the teaching in your class?	Never – very often (5-point scale)

Note: *)Items 9a, 9b, and 9c are related to two factors each, *Interest* and *SpSw*, *SpEn* and *SpMa*, respectively.

differences, with the results showing that systematic differences existed between schools, where intra-class correlations for the variables ranged from .062 to .085 for subject grades and national tests in Swedish, English, and mathematics. In order to take account of clustering effects, the ''complex'' option offered by the Mplus program (Muthén & Muthén, 2004) was used in the current study. Whilst this method compensates for disturbances in the χ^2 and standard errors due to clustering effects, it does not affect the estimates. Since the differences in the estimates on the student and school level for subject grades and national tests were small (Klapp Lekholm & Cliffordson, 2008) and the intra-class correlations for the student characteristics included in this study also were very small, there is no need to conduct a full two-level analysis. In the ''complex'' analyses, the standard errors become larger and the *t* values become smaller due to losses in information caused by the clustering. The extent of the information loss due to clustering effects is a function of the intra-class correlation and the cluster size (Muthén & Muthén, 2004).

The starting point for the current study was the result from a previous study (Klapp Lekholm & Cliffordson, 2008), where a baseline four-factor model (model A) with four latent variables (*Sw*, *En*, *Ma*, *ComGr*) was developed to reflect both the three subject-specific or achievement factors and a common grade factor that cuts across the different subjects (Swedish, English, and mathematics). The subject-specific factors (*Sw*, *En*, *Ma*) were related to the grades to each subject (SGSW, SGEN, SGMA) and for each of the

national test scores (NTSW1-3, NTEN1-3, NTMA) in Swedish, English, and mathematics, respectively, with covariances between *Sw*, *En* and *Ma*. In order to separate the common grade dimension, the *ComGr* factor was only related to the three subject grades in Swedish, English, and mathematics (SGSW, SGEN, SGMA).

The next step involved the estimation of a measurement model (Model B) with the eight student characteristic factors related to all of their respective indicators, with covariances between the factors.

In order to investigate the relations between the different student characteristics and the subject-specific and the common grade dimensions, respectively, the next step was to add the student characteristic model (Model B) to the previously estimated baseline 4-factor model (Model A) and create a 12-factor model (Model C) with covariances between the eight student characteristic factors (*Parent*, *Interest*, *Coping*, *Adjust*, *Cooperate*, *SpSw*, *SpEn*, *SpMa*) and the three subject-specific factors (*Sw*, *En*, *Ma*), and between the eight student characteristic factors and the common grade factor (*ComGr*), respectively.

Since the previous study (Klapp Lekholm & Cliffordson, 2008) had found gender differences in the subject-specific and the common grade dimensions, the next step in the modelling process was to estimate the previous baseline four-factor (*Sw*, *En*, *Ma*, *ComGr*) model (Model A) with relations between gender and the four factors (Model D) (see Figure 1). A dummy variable for gender (0 = male, 1 = female) was related to the four factors.

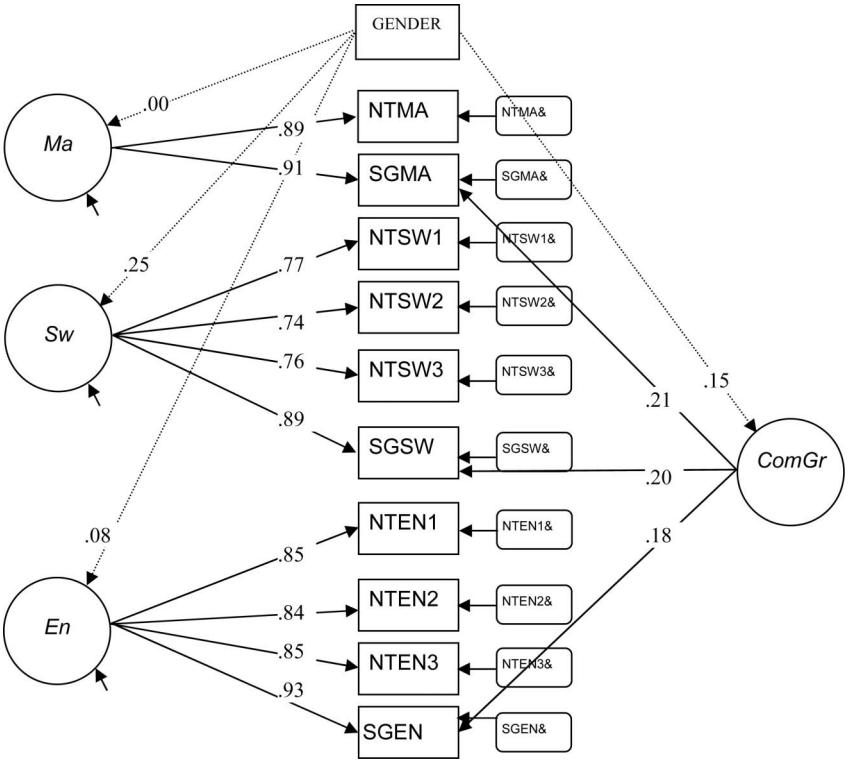

Figure 1. Gender as dummy variable related to all the factors: achievement in Swedish, English, and mathematics (*Sw*, *En*, *Ma*) and to the common grade dimension (*ComGr*), and with covariances between the residuals for *Sw*, *En*, and *Ma*. The indicators are national tests and subject grades in each subject (NTSW1-3, NTEN1-3, NTMA, SGSW, SGEN, SGMA) (Model D).

In order to investigate and explain the mediating effects of gender on grades, the next step in the modelling process was to add the student characteristic factors (*Parent*, *Interest*, *Coping*, *Adjust*, *Cooperate*, *SpSw*, *SpEn*, and *SpMa*) to Model D. They were related to the subject-specific factors *Sw*, *En*, *Ma* and the common grade factor, *ComGr*, one at a time, and hence were estimated eight times (Models E1–E8) (see Figure 2).

In order to estimate the amount of school differences in the different student characteristics intra-class correlations were computed. The baseline four-factor model (Model A) was estimated with one student characteristic at a time and with covariances between all the manifest variables on the school level.

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 the number of free parameters into account (Jöreskog, 1993). The RMSEA should be below .08 for a model to be acceptable, whereas to be good, the RMSEA should be below .05.

The data used for grades and national tests include the whole population, whereas the questionnaire data comprise a 10% nationally representative sample. The missing information was handled using missing data modelling (Muthén, Kaplan, & Hollis, 1987), which, under relatively moderate assumptions, produces unbiased estimates (Schafer & Graham, 2002). The missing data analysis applied makes the assumption

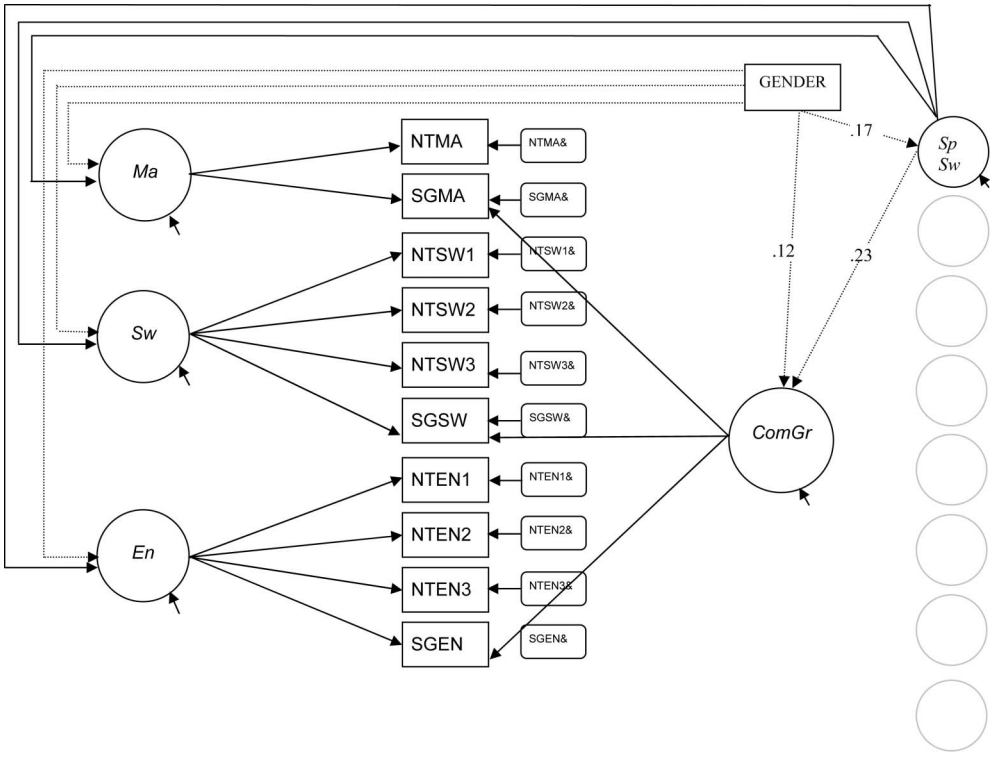

Figure 2. The structural models (Models E1–8) with relations between one characteristic factor at a time and the subject-specific and common grade factors (*Sw*, *En*, *Ma*, and *ComGr*) and with covariances between the residuals for *Sw*, *En*, and *Ma*. Gender is a dummy variable with a direct and a mediating effect on *Sw*, *En*, *Ma*, and *ComGr*.
Note: The estimates given are an example of the direct and indirect relations in model E1.

that the data are "missing at random" (MAR), which implies that the procedure 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 are "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 purposes of estimation and for testing all of 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, gender, and questionnaire data are shown in Table 2.

The descriptive statistics showed that there were missing observations for some of the variables, particularly in the case of the national test in mathematics, where there was a considerable proportion of missing data (43.1%). 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 content of the test was leaked in advance. There is, however, no reason to assume that this missingness is biased due to achievement in mathematics. For the national tests in Swedish and English, the proportion of missing data ranged between items from 17.8% for NTEN2 to 21.3% for NTSW2. For the subject grades Swedish, English, and mathematics, the missing data proportion was low, 0.7%. The missingness in the questionnaire data ranged from 26.4 to 27.9%.

Table 2. Descriptive statistics for the manifest variables; subject grades, national tests, gender, and questionnaire data.

Variables	N	% missing	M	SD
Population				
Subject Grades				
SGSW	98 353	0.7	13.0	4.3
SGEN	98 353	0.7	13.1	4.7
SGMA	98 353	0.7	12.1	4.5
National tests				
NTSW1	81 391	17.8	11.8	4.8
NTSW2	77 979	21.3	12.9	4.0
NTSW3	81 131	18.1	12.2	4.1
NTEN1	79 832	19.4	13.2	4.3
NTEN2	81 426	17.8	13.8	5.0
NTEN3	81 221	18.0	12.7	4.4
NTMA	56 325	43.1	11.9	4.4
Gender				
Girls	48 660	0.0		
Boys	50 410	0.0		
Sample				
Questionnaire data				
43 items	6 289–6 416	26.4–27.9		

Note: The questionnaire data contains missing on every item to fairly the same amount.

The baseline four-factor model (Model A)

The first model was the previously estimated model (Klapp Lekholm & Cliffordson, 2008) with four factors (*Sw*, *En*, *Ma*, *ComGr*) where the subject-specific factors were specified by subject grades and national tests in each subject (SGSW, SGEN, SGMA, NTSW1-3, NTEN1-3, and NTMA) and the common grade factor (*ComGr*) was specified by subject grades in each subject, and with covariance between *Sw*, *En*, and *Ma*. The goodness-of-fit of this model was good ($\chi^2(29, 99070) = 5061.25$; RMSEA = .042). The standardized factor loadings showed that the subject-specific factors (*Sw*, *En*, *Ma*) related substantially to the subject grades (.90 to .93) and to the national tests (.74 to .89). The *ComGr* factor was related to all subject grades, the factor loadings being .23 for Swedish, .19 for Mathematics, and .17 for English. This result shows that grades are multidimensional and that it is possible to separate two dimensions in grades; a subject-specific and a common grade dimension. The factor loadings for the *ComGr* factor may seem relatively small, but the effects exist across all subjects, which makes their contribution to the sum more important than is indicated by the proportion of variance accounted for in each grade (Reuterberg & Gustafsson, 1992).

The student characteristic model (Model B)

The first step was to estimate the student characteristic measurement model with eight factors (*SpSw*, *SpEn*, *SpMa*, *Parent*, *Coping*, *Interest*, *Adjust*, and *Cooperate*) with relations to all their respective indicators from the questionnaire data and with covariances between the factors (see Table 1). This model validates the factors and suggests that the indicators were reasonable. All the standardized factor loadings of this model were significant on the .001 level. The fit indices were acceptable (χ^2 (828, 8717) = 14995.78; RMSEA = .051). Table 3 presents the standardized factor loadings for this model, which show that the *SpSw*, *SpEn*, and *SpMa* factors accounted for variance in all their respective indicators. Some items, for example, ''interest in learning more in Swedish'' (.36) and ''read subtitles'' (.39), had relatively low loadings but were nevertheless significant. The *SpEn* and *SpMa* factors had generally higher loadings in comparison to the *SpSw* factor. The factor loadings for the rest of the factors (*Parent*, *Coping*, *Interest*, *Adjust*, and *Cooperate*) also showed a pattern where a few items, for example, ''if parents participate in school'' (.29) and ''are you considering going to the university'' (.20), had relatively low loadings but were nevertheless significant and theoretically reasonable.

Student characteristics added to the baseline four-factor model (Model C)

The next step in the modelling process was to combine Model A and Model B and thus to estimate a 12-factor model (*Sw*, *En*, *Ma*, *ComGr*, *SpSw*, *SpEn*, *SpMa*, *Parent*, *Interest*, *Coping*, *Adjust*, *Cooperate*) with covariances, on the one hand, between the eight student characteristic factors and the common grade factor (*ComGr*) and, on the other hand, between the eight student characteristic factors and the subject-specific factors (*Sw*, *En*, *Ma*) and also between the residuals of *Sw*, *En*, and *Ma*. The goodness-of-fit indices were excellent (χ^2 (1255, 99070) = 21773.91; RMSEA = .013). See Table 4 for an overview of the covariances.

The relations between the three *SpSw*, *SpEn*, *SpMa* factors and their respective subject-specific factors, *Sw*, *En*, and *Ma*, (.63, .70, .65, respectively) were substantial. The relation between the *SpSw* factor and the *En* factor (.55) also was quite considerable.

Table 3. Standardized factor loadings for student characteristic, eight-factor model (B) with covariances between the factors.

Latent variables Manifest variables	*SpSw*	*SpEn*	*SpMa*	*Parent*	*Coping*	*Interest*	*Adjust*	*Cooperate*
1a. READ	.65							
1b. REA_AL	.68							
1c. REA_SU	.39							
1d. WRITE	.59							
1e. DISSCU	.62							
1f. PRESEN	.64							
2. GOOD_S	.65							
9a. INTE_S	.36							
3a. UND_TA		.76						
3b. UND_FI		.73						
3c. READ		.80						
3d. TALK		.78						
3e. WRITE		.78						
3f. DICTIO		.70						
3g. PRESEN		.72						
4. GOOD_E		.79						
9b. INTE_E		.43						
5a. CALCUL			.70					
5b. ADD_DIV			.68					
5c. PERCEN			.74					
5d. ARE_CI			.72					
5e. EQUATI			.76					
5f. SO_PRO			.80					
5g. EX_PRO			.71					
6. GOOD_M			.81					
9c. INTE_M			.50					
7a. FEEL				.78				
7b. WORK				.89				
7c. SHOW				.64				
7d. PART				.29				
8a. KEEP_U					.69			
8b. GIVE_U					.72			
8c. HELP					.67			
8d. CONCEN					.67			
9a. INTE_S						.67		
9b. INTE_E						.62		
9c. INTE_M						.48		
10. UNIVER						.20		
11a. CLASS							.61	
11b. SCHOOL							.74	
11c. PUPILS							.64	
11d. TEACHE							.70	
11e. WORK							.69	
12a. CLA_DIS								.58
12b. GROU_W								.39
12c. PU_PLA								.58

Note: All estimates are significant on the .001 level. For a closer description of the variables see Table 1.

Table 4. Standardized covariances for relations between student characteristic and the four-factor model (C) with covariances between the *Sw*, *En*, *Ma* factors and the student characteristic factors, and between the *ComGr* factor and the student characteristic factors.

Latent variable	*Sw*	*En*	*Ma*	*Com Gr*	*Sp Sw*	*Sp En*	*Sp Ma*	*Parent*	*Coping*	*Interest*	*Adjust*
Sw											
En	.79										
Ma	.70	.66									
SpSw	63	.55	.36	−.08							
SpEn	.46	.70	.33	−.12	.70						
SpMa	.33	.28	.65	.09	.39	.28					
Parent	.25	.16	.16	.12	.33	.20	.22				
Coping	.42	.41	.48	.13	.46	.41	.55	.24			
Interest	.08	−.04^{ns}	−.02^{ns}	.26	.07^{ns}	−.09	.01^{ns}	.22	.04^{ns}		
Adjust	.14	.14	.18	.14	.29	.21	.34	.24	.52	.16	
Cooperate	.10	.09	.07	.01^{ns}	.23	.17	.19	.28	.22	.16	.45

Note: ^{ns}All estimates are significant on the .001 level except for the marked values, which are nonsignificant.

The relations between the other characteristic factors, *Parent*, *Coping*, *Interest*, *Adjust*, *Cooperate* and the *Sw*, *En*, and *Ma* factors were all positive, except for the relation between the *Interest* factor and the *En* and *Ma* factors, which were nonsignificant. The result of this model also suggests that *Parent* has a higher influence in the *Sw* factor (.25) compared to the *En* and *Ma* factors (.16). The *Coping* factor seems to be of great importance in the *Sw*, *En*, and *Ma* factors (.42, .41, .48, respectively).

The relations for the different characteristic factors and the common grade factor indicated that the *SpSw* and *SpEn* factors were significantly negatively related to *ComGr* (–.08 and –.12, respectively), while the *SpMa* factor was significantly positively related to *ComGr* (.09). The *Interest* factor had the strongest relation to *ComGr* (.26). Three other factors, *Parent, Coping*, and *Adjust* were also significantly positively related to the common grade factor (.12, .13, .14), whilst the *Cooperate* factor had a small and nonsignificant relation with *ComGr*.

So far, the results show that the *ComGr* factor explains variance in grades and that certain student characteristics are of importance. Student motivation (*Interest*) influences *ComGr* the most, while student characteristics such as adjustment (*Adjust*), coping (*Coping*), and parental engagement (*Parent*) influence the *ComGr* factor to essentially the same degree.

The baseline four-factor model with gender included (Model D)

Since previous research has shown that there are gender differences in grades, gender was added to the baseline model (Model D; Figure 1). The standardized factor loadings for this model were substantial, and the goodness-of-fit indices were better (χ^2 (35, 99070) = 4494.06; RMSEA = .036) than for Model A. The result indicated that there was a gender difference in the *ComGr* factor of .15, with an advantage for girls. The result also showed that gender was related to the *Sw* and *En* factors (.25 and .08, respectively) but not to the *Ma* factor (0.00).

Student characteristics and the baseline four-factor model with direct and indirect influence of gender (E1–E8)

In order to investigate the direct and indirect effect of gender on the common grade dimension, the student characteristic factors were added to the baseline four-factor model with gender as a dummy variable related to all four factors in the model (Model D), (*Sw*, *En*, *Ma*, *ComGr*, *SpSw*, *SpEn*, *SpMa*, *Parent*, *Coping*, *Interest*, *Adjust*, *Cooperate*) each individually in succession (see Figure 2). This resulted in eight models (E1–E8) estimated with the same relations but with a different student characteristic factor. The goodness-of-fit indices were good for all the models. Table 5 presents the standardized factor loadings and goodness-of-fit indices.

In this model, the direct effect of gender on the *ComGr* factor, as well as the indirect effect of gender, mediated through the characteristics factors on the *ComGr* factor, was estimated. Here an example is given in order to clarify the models presented in Table 5. For example (see Figure 2), the relation from the student characteristic factor *SpSw* to the *ComGr* factor is .23 and the direct relation from GENDER to *ComGr* is .12 (in Model D, this relation was .15). The indirect relation goes from the GENDER variable to the student characteristic factor *SpSw* and then to the *ComGr* factor, and the GENDER and the *SpSw* factor has a relation of .17. Since the direct relation is somewhat smaller in this model (.12), this result suggests that the student characteristic S*pSw* mediates some of the gender difference found in Model D.

The *Interest* factor explained almost all of the gender differences in the *ComGr* factor, while the *Parent* factor explained some gender differences in *ComGr*. The models with *Coping*, *Adjust*, and *Cooperate* factors, mediated no gender effect on the *ComGr* factor.

The direct and indirect effect of gender on the subject-specific factors influenced only the *Sw* and *En* factors, where *Interest* mediated a large amount of the gender differences in these factors, while the *Parent* factor mediated some of the gender differences.

Table 5. Relations for the models, E1–E8, with the direct and indirect influence of gender on factors.

Model	Latent variable	*Sw*	*En*	*Ma*	*ComGr*	GENDER	χ^2 *df*	RMSEA
E1	*SpSw*	.21			.23	.17	9791.73	.026
	GENDER	.22	.08	.00^{ns}	.12		140	
E2	*SpEn*		.37		−.25	−.02^{ns}	9865.21	.025
	GENDER	.26	.09	.00^{ns}	.14		158	
E3	*SpMa*			.42	.31	−.11	8296.44	.023
	GENDER	.25	.08	.05	.18		158	
E4	*Parent*	.20	.08	.18	.31	.19	5906.63	.028
	GENDER	.22	.06	−.04	.10		76	
E5	*Coping*	.44	.34	.48	.39	−.11	5478.35	.027
	GENDER	.30	.11	.05	.21		76	
E6	*Interest*	.38	.26	.33	.39	.27	7381.86	.031
	GENDER	.16	.01^{ns}	−.09	.06		76	
E7	*Adjust*	.13	.03^{ns}	.17	.38	−.08	6282.00	.026
	GENDER	.27	.08	.01	.18		90	
E8	*Cooperate*	.08	−.03^{ns}	.10	.37	−.02^{ns}	5425.47	.029
	GENDER	.26	.08	.00^{ns}	.16		63	

Note: ^{ns}All values are significant on the .001 level except for the marked values, which are nonsignificant.

In order to estimate the amount of school differences in the different student characteristics, intra-class correlations were computed. The baseline four-factor model (Model A) was estimated with one student characteristic at a time and with covariances between all the manifest variables on the school level. The models with the different student characteristics all showed a good fit where, for example, the fit indices for the model with student interest included were χ^2 (64, 99070) = 5637.73; RMSEA = .030). The results showed that the intra-class correlations for the variables for national tests and grades were in line with the previous study, which is to say between .06 and .09 (Klapp Lekholm & Cliffordson, 2008). No more than 2.6 to 5.3% of the variance in the items concerning student general interest or motivation was explained by school differences. In the model with the *Parent* factor 0.9 to 3.3% of the variance in the items for parental support was due to school differences. That is, the overall pattern was that the student characteristic variables accounted for relatively lower variance on the school level. When explained variance is below 5%, it is not reasonable to believe that systematic differences due to school effects are of major importance (Hox, 2002). This result seems reasonable, since individual characteristics vary primarily on the individual level.

Conclusions and discussion

The purpose of this study was to investigate which different student characteristics influence grades. Another purpose was to investigate the influence of gender on grades and how different student characteristics mediate the effect of gender on grades. Previous findings (Klapp Lekholm & Cliffordson, 2008) identified and separated two dimensions in grades: the subject-specific (cognitive factors) and the common grade dimension (noncognitive factors) which have been the starting point of this study. The primary focus has been to investigate and explain variance related to student characteristics in the common grade dimension in grades.

The results show that student characteristics influence grades and that some student characteristics primarily influence the common grade dimension, while other student characteristics primarily influence the subject-specific dimension. Student motivation influences the common grade dimension in grades the most, while student self-efficacy and self-perceived competence in the three subjects influence the subject-specific dimension in grades the most. The results also show that student general interest or motivation mediates almost all the influence of gender in the common grade dimension.

Which student characteristics are related to grades?

Students' self-perception of competence is strongly correlated with the subject-specific dimension in grades. Positive self-perception of competence seems to be a good predictor of achievement, which is in line with previous findings (Brookhart & Devoge, 1999; Brookhart & Durkin, 2003). However, the students in this study have been awarded grades on three previous occasions during compulsory school, which, of course, has affected their self-perception of competence in the different subjects. Self-perception of competence may also be affected by the classmates' performance levels. It is thus not possible, based on the results from this study, to argue that students' self-perception of competence causes achievement or high grades.

Students' self-efficacy or coping strategies influence the subject-specific dimension in grades quite substantially, which is in agreement with previous research (Brookhart & Devoge, 1999, Brookhart & Durkin, 2003). It seems as if the way students judge their

efficacy is a strong predictor of student achievement. Parental engagement in school work and whether students are happy in school also seem to influence the subject-specific dimension in grades, but not to the same amount as self-efficacy. That students cooperate with their peers and teachers and participate in planning their education also seem to be characteristics that influence the subject-specific dimension in grades. Wentzel (1991) found that socially responsible behaviour, for example, cooperation, was highly relevant for academic success, a finding which is in line with the results of the current study, namely that cooperation is of importance for the subject-specific dimension and achievement in certain subjects.

However, some of these relations are different for the common grade dimension. The fact that the common grade dimension cuts across the different subjects suggests that this dimension represents factors with a broad range of influence. The results show that student motivation is positively related to the common grade dimension. Students' specific interests (*SpSw*, *SpEn*, *SpMa*) influence the subject-specific dimension, whereas students' general interests or motivation are of importance primarily for the common grade dimension. It has also been stressed that interest and motivation are closely related and that both individual and situational interests are motivational (Brookhart et al., 2006). The relation between general interest or motivation and the common grade dimension seems to recognize that students who are motivated often possess both specific and general goals and approach new phenomena with the goal of understanding them, which is a student characteristic awarded in grades. The importance of student motivation for grade outcomes found in this study indicates that this is a fundamental student characteristic that teachers value when grading (Hidi & Renninger, 2006; Hidi et al. 2004). The result also explains the relation between student motivation and the subject-specific and common grade dimensions and shows that general interest or motivation is of importance primarily in the common grade dimension and of less importance in the subject-specific dimension. This implies that student motivation cuts across the different subjects.

The common grade dimension also relates positively to student self-perceived competence in mathematics (*SpMa*) but negatively to student self-perceived competence in Swedish and English (*SpSw*, *SpEn*). One explanation may be that students' self-perceived competence in the different subjects corresponds to student achievement, which explains why the common grade dimension is not related to self-perceived competence in Swedish and English but related to self-perceived competence in mathematics.

How does gender influence grades?

The result shows that gender differences exist primarily in the Swedish and English subject-specific dimension and in the common grade dimension, which suggests that females are favoured in these dimensions and are awarded higher grades in comparison to their results on the national test. This result supports previous research that has indicated that females are awarded higher grades in comparison to their result on national tests (Andersson, 1998; National Agency for Education, 2007b). One explanation is that girls perform, based on expectations from others and own experiences (Gipps & Murphy, 1994; Murphy, 2000), better in the dimension related to languages (Swedish and English) and develop a more general interest or motivation towards learning, which is favoured in the common grade dimension. The female advantage in the common grade dimension is explained by girls' greater motivation for learning, which cuts across the different subjects.

The results show that student motivation, parental engagement in school, and self-perception in Swedish all mediate gender differences on the common grade dimension.

Students' motivation to learn appears to mediate almost entirely the gender difference in the common grade dimension. Thus, one important mechanism behind the advantage for girls in grades seems to be that they express a greater degree of interest and motivation in schoolwork than boys. According to research, student general interest and motivation are of major importance for student achievement in school, which explains why girls are awarder higher grades in comparison to their test results and to boys. The different approaches towards learning adopted by boys and girls may provide one possible explanation, in that boys tend to nurture their specific cognitive abilities whilst girls tend to nurture their general cognitive abilities (Rosén, 1998). Another explanation for the influence of student motivation which mediates the gender difference in the common grade dimension suggests that girls have developed more social skills than boys and are better prepared for school work and perform better on the broad ability dimension which is rewarded in school and in grades (Ainley et al., 2002; Arnot et al., 1996; Bergin, 1999; Murphy, 2000).

The identification of mediating factors can provide an explanation for the observed gender differences in grades, in terms of greater interest and motivation and stronger parental engagement in school work. If girls are better prepared for schooling, talk about school at home with parents, this may affect the grades over and above the results achieved on tests. In families where parents and children talk about school, school work, and how the schoolwork is progressing may represent family values which are favoured in school. The experiences that girls and boys face during their upbringing thus affect their chances in school in that education assumes that students should demonstrate their knowledge primarily via reading and writing, as well as through social behaviour such as being interested in learning and, in addition, having family values that are in line with school values (Murphy, 2000). These noncognitive and social factors seem to be of importance in understanding gender differences in achievement and grades.

Construct-irrelevant variance in grades

The complex issue about the relevance of different student characteristics in grades and, in the long term, whether grades are fair, concerns how we define construct-irrelevant variance in grades (Messick, 1994). Which student characteristics are relevant in grades and should be a part of grades concerns difficult issues about construct validity and its consequential aspects in particular. Society needs students who are motivated and persistent in their pursuit of study goals, and it is these students who are rewarded with success both in grades in the educational system and in society at large.

In a highly decentralized educational system, such as in Sweden, the question about the comparability of grades between students and between educational systems is frequently raised. However, teachers seem to assess and grade students in a manner which is not completely in line with the curriculum in the sense that their grading accords partly with what is expected of students from society at large. One possible explanation may be that teachers are well aware of the demands from society on student abilities, both cognitive and noncognitive, and that this "teacher knowledge" is not recognized formally either in the curriculum or in the educational system. Teachers find themselves in an intersection between their own beliefs and values in the domain of learning, instruction, and the awarding of grades and the external pressure and demands from curriculum and steering documents, as well as demands from society at large (Deutsch, 1979; Pilcher, 1994; Wiliam, 1996). The results from this study place an emphasis on issues about the relevance of the different kinds of student characteristics that are to be assessed and whether a

student characteristic such as motivation is of more importance for students' overall educational and working success than what is recognized in the curriculum.

Limitations

One limitation of the current study is that the number of investigated student characteristics is somewhat limited and that there are several other student characteristics which could affect grades that are of potential interest, such as students' ambitions and plans for the future. The issue of causality is also difficult to reveal in this study. Another limitation concerns the use of the national tests as instruments for measuring achievement. The national tests are assumed to measure achievement or subject-specific knowledge, but it seems reasonable to believe that the tests also measure some amount of non-achievement, since the national tests are administered and marked locally by the teachers themselves.

Yet another limitation of the current study is that observations were not available on all variables for all persons, so missing-data modeling had to be relied upon. While the maximum likelihood missing-data procedure used relies on relatively mild assumptions, and the large sample size should assure its asymptotic properties (Schafer & Graham, 2002, p. 164), it may be noted that the proportion of cases with complete data was relatively low. This may warrant some caution in interpreting the results from the current study.

Future research

Further analyses of the current data should be conducted in order to reveal school differences in the common grade dimension and in grades. School characteristics, such as type of school, location, and the characteristics of the teaching staff, can be investigated more closely in order to understand how systematic differences in these factors influence grades.

Acknowledgements

The Swedish Research Council has financially supported the research reported in this paper. The work is a part of the GRAM (Grades and grade assignment: Functions and effects) project.

Notes on contributors

Alli Klapp Lekholm is a PhD student at Gothenburg University, Sweden. She is working with research on grades and grade assignment within the GRAM (Grades and grade assignment: Functions and effects) project.

Christina Cliffordson, associated professor in education, is working with research on assessment and measurement based on large-scale longitudinal data by the use of multivariate analyses methods.

References

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

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *3*, 545–561.

Alexander, W.P. (1935). Intelligence, concrete and abstract. *British Journal of Psychology Monograph Supplement*, *19*, 177.

Andersson, A. (1998). The dimensionality of the leaving certificate in the Swedish compulsory school. *Scandinavian Journal of Educational Research*, *42*(1), 25–40.
Arnot, M., David, M., & Weiner, G. (1996). *Educational reforms and gender equality in schools*. Manchester, UK: EOC.
Bergin, D.A. (1999). Influences on classroom interest. *Educational Psychologist*, *34*, 87–98.
Brookhart, S.M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, *10*(1), 35–36.
Brookhart, S.M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*(2), 123–142.
Brookhart, S.M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, *7*(4), 279–301.
Brookhart, S.M., & Devoge, J.G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, *12*(4), 409–425.
Brookhart, S.M., & Durkin, D.T. (2003). Classroom assessment, student motivation and achievement in high school social studies classes. *Applied Measurement in Education*, *16*(1), 27–54.
Brookhart, S.M., Walsh, J.M., & Zientarski, W.A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education*, *19*(2), 151–184.
Cliffordson, C. (2004). Betygsinflation i de målrelaterade gymnasiebetygen [Inflation in goal-referenced grades from upper secondary school]. *Pedagogisk forskning i Sverige*, *9*(1), 1–14.
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.
Deutsch, M. (1979). Education and distributive justice. *American Psychologists*, *34*(5), 391–401.
Giota, J. (2006). Why am I in school? Relationships between adolescents' goal orientation, academic achievement and self-evaluation. *Scandinavian Journal of Educational Research*, *4*, 441–461.
Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckinghamshire, UK: Open University Press.
Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioural Research*, *28*(4), 407–434.
Gustafsson, J.-E., & Stahl, P.-A. (2005). *STREAMS 3.0 user's guide*. Mölndal, Sweden: Multivariateware.
Härnqvist, K. (2000). Evaluation through follow-up. A longitudinal program for studying education and career development. In C.-G. Janson (Ed.), *Seven Swedish longitudinal studies in behavioral science*. Stockholm: Forskningsrådsnämnden.
Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 47–77.
Hidi, S., & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist*, *4*, 111–127.
Hidi, S., Renninger, K.A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D.Y. Dai & R.J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 89–115). Mahwah, NJ: Lawrence Erlbaum Associates.
Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
Jöreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.
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*, 181–199.
Lane, K.L., Givner, C.C., & Pierson, M. (2004). Secondary teacherś view on social competence: Skills essential for success. *The Journal of Special Education*, *3*, 174–186.
McMillan, J.H. (2003). The relationship between instructional and classroom assessment practices of elementary teachers and student scores on high-stakes tests. Manuscript submitted for publication in *ERIC, Clearinghouse on Assessment and Evaluation*.
Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, *23*(2), 13–23.

Murphy, P. (2000). Equity, assessment and gender. In J. Salisbury & S. Riddell (Eds.), *Gender, policy and educational change* (pp. 134–152). London: Routledge.
Murphy, P., & Whitelegg, E. (2006). Girls and physics: Continuing barriers to belonging. *Curriculum Journal*, *3*, 281–305.
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 ed.). Los Angeles: Authors.
National Agency for Education (2003). *Skolverkets lägesbedömning 2003* [The National Agency for Educatioń s determination of position 2003]. Stockholm: National Agency for Education.
National Agency for Education (2005). *Skolverkets lägesbedömning 2005* [The National Agency for educatioń s determination of position 2005]. Stockholm: National Agency for Education.
National Agency for Education (2007a). *National subject tests in compulsory school*. Retrieved May 2, 2007, from http://www.skolverket.se/sb/d/276
National Agency for Education (2007b). *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 (2007c). *Steering documents*. Retrieved October 2, 2007, from http://www.skolverket.se/sb/d/287
Pilcher, J.K. (1994). The value-driven meaning of grades. *Educational Assessment*, *2*(1), 69–88.
Pintrich, P.R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, *41*, 219–225.
Reuterberg, S.-E., & Gustafsson, J.-E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement*, *52*(4), 795–811.
Rosén, M. (1998). *Gender differences in patterns of knowledge* (Göteborg studies in educational science, Vol. 124). Göteborg, Sweden: Acta Universitatis Gothoburgensis.
Schafer, L.J., & Graham, W.J. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.
Wang, J., & Lin, E. (2005). Comparative studies on U.S. and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher*, *5*, 3–13.
Wentzel, K.R. (1989). Adolescent classroom goal, standards for performance, and academic achievement: An interactionist perspective on primary prevention. *Journal of Consulting and Clinical Psychology*, *59*, 830–841.
Wentzel, K.R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, *62*, 1066–1078.
Wikström, C. (2005). *Criterion-referenced measurement for educational evaluation and selection. No. 1*. Umeå, Sweden: 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.