

# Does academic and social self-concept and motivation explain the effect of grading on students' achievement?

Alli Klapp<sup>1</sup> 

Received: 21 June 2016 / Revised: 24 February 2017 / Accepted: 27 February 2017 /

Published online: 10 March 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** The purpose of the study was to investigate if academic and social self-concept and motivation to improve in academic school subjects mediated the negative effect of summative assessment (grades) for low-ability students' achievement in compulsory school. In two previous studies, summative assessment (grading) was found to have a differentiating effect on students' subsequent achievement in school (Klapp et al. 2014; Klapp 2015). Differences between subgroups of students (cognitive ability, gender, and socioeconomic status) were controlled for in the analyses. Data was retrieved from the Evaluation Through Follow-up (ETF) longitudinal project containing register and questionnaire data on a large national representative sample of Swedish compulsory school students ( $N = 8558$ ). Due to a unique natural circumstance, municipalities in Sweden could decide whether or not to grade their students in sixth grade which made it possible to apply a quasi-experimental design. In the sample, 50% of the students were graded in sixth grade while all students were graded in seventh grade. Confirmatory factor analysis (CFA) and structural equation modeling (SEM) have been estimated. The results show that the negative effect of summative assessment (grading) for low-ability students on their subsequent achievement is fully mediated by academic self-concept in mathematics and Swedish, and motivation to improve in academic school subjects.

**Keywords** Academic self-concept · Achievement · Grades · Compulsory school · Longitudinal study · Cognitive ability · Gender

---

✉ Alli Klapp  
alli.klapp@ped.gu.se; <http://www.ips.gu.se>

<sup>1</sup> Department of Education and Special Education, University of Gothenburg, Box 300, SE-405 30 Gothenburg, Sweden

## Introduction

Researchers, policymakers, and society at large seem to agree that summative assessments such as grading and testing affect students' motivation for learning and achievement in school. In several reviews on the impact of summative assessments on students' learning, motivation, and achievement, it has been shown that low-ability students are negatively affected by high-stake testing and grading (Harlen and Deakin Crick 2002; Natriello 1987), some students experience test anxiety (Harlen and Deakin Crick 2002), and in many cases, students would learn more if not being assessed in high-stake summative environments (Black and Wiliam 1998; Crooks 1988; Kluger and DeNisi 1996).

One possible explanation for the found negative effect of summative assessments such as grading on low-ability students' achievement may be that these students have a lower cognitive ability. However, even high-ability students have been shown to be affected negatively by high-stake assessment regimes (Carrillo-de-la-Peña et al. 2009; Cilliers et al. 2010; Trotter 2006) and are favored by less summative assessments and being assessed after learning has taken place. Another possible explanation for the negative grading effect may be that students' non-cognitive or affective skills such as self-concept and motivation are affected by summative assessment regimes (Wentzel et al. 2010) and thus affect their learning and achievement in school (Harlen and Deakin Crick 2002).

In a recent study, Klapp et al. (2014) used a large national representative sample of students in sixth grade in compulsory school and compared students' subsequent achievement in seventh grade for ungraded students and for students who received grades at the end of the school year in sixth grade. In Sweden, grades are used for summarizing students' achievement at the end of a semester or school year and with the purpose to summarize and to be used as an instrument for selection. This investigation was possible due to a unique circumstance in the Swedish educational system at the beginning of the 1980s where municipalities had a choice whether or not to grade their students in sixth grade. This quasi experiment combined with a national longitudinal research project Evaluation Through Follow-up (ETF) (see Härnqvist 2000), national representative data for a large sample of students were gathered by Statistics Sweden and are available for research. About 50% of the students in the ETF sample were graded in sixth grade, while 50% were not. Klapp et al. (2014) found that grading had differential effects due to students' ability and gender. Students with low cognitive ability and boys were negatively affected in their subsequent achievement by grading, compared to students with low cognitive ability who were not graded and girls. In a follow-up study, Klapp (2015) found negative effects of grading on achievement up through compulsory school and graduation from upper secondary school for low-ability students and boys, compared to ungraded low-ability students and girls. In the ETF project, register data and student self-reports are available including measures of different aspects of students' academic and social self-concept and motivation (motivation to improve in academic school subjects). Thus, the current study differs from these two previous studies by including a student self-report which measures academic and social self-concept and an aspect of motivation defined as the motivation to improve in academic school subjects.

By adding new data to the previous models and by using modern powerful statistical methods, this unique dataset can be analyzed in order to investigate the mediating relations between students' academic and social self-concept and motivation to improve in academic school subjects for graded and ungraded students' subsequent achievement.

## Previous research

### The impact of summative assessments on students' motivation for learning and achievement

Harlen and Deakin Crick (2002) concluded in a systematic research review that summative assessments have differentiating effects on students' motivation for learning and achievement. They found that low-achieving students were negatively affected in their motivation for learning, by tests and grades, compared to high-achieving students. When high-stake tests were introduced in England, a negative correlation between self-esteem and achievement became evident, contrary to before the introduction of the high-stake testing regime, where such correlation did not exist. Besides, they found that frequent testing reinforced low self-image of the low-achieving students. One reason why summative assessments differentiate students was suggested to be due to the association between summative assessments and students' personal characteristics, background, and prerequisites (Alexander 2010; Harlen and Deakin Crick 2002). Besides, high achievers seem to be less negatively affected by grading than low achievers (Pollard et al. 2000; Klapp et al. 2014; Klapp 2015) and high achievers seem to have a better understanding of grades and are less affected by the type of feedback used by the teacher (Butler 1988).

These effects of a summative assessment environment on students' self-concept and motivation to improve seem to be different due to gender (see, for example, Eccles 1984; Conger and Long 2010; Herbert and Stipek 2005; Jansen et al. 2014) and cognitive ability (see, for example, Dweck and Yeager 2012; Dweck and Leggett 1988; Nicholls and Miller 1984). Thus, the long-term negative effect of summative assessments for low-ability students' achievement may partly be due to their lower cognitive ability, to background characteristics, and to their lower self-concept and motivation for improving their academic skills (motivation to avoid failure) which may affect their learning process and achievement.

Some studies, in contrast, have reached the conclusion that summative assessments influence students' learning and achievement positively (Artes and Rahona 2013; Becker and Rosen 1992; Bandiera et al. 2008). When giving students the possibility to make social comparisons within the classroom, they achieve higher (Azmat and Iriberry 2009), and competition between students stimulates their academic effort (Becker and Rosen 1992), and girls and students with parents with a low socioeconomic status (SES) are advantaged in their achievements if being graded in compulsory school (Sjögren 2010).

### The relations between academic and social self-concept and motivation to improve in academic school subjects and students' achievement

#### *The importance of academic and social self-concept for achievement in school*

Self-concept has consistently been shown to be of importance for student's achievement in school (Caprara et al. 2008; Wentzel 1991). Self-concept is broadly defined as a person's perception of herself or himself, and this perception is formed through personal experiences and interpretations with and of the environment and significant others (Shavelson et al. 1976). The hierarchical construct is structured into a broad general self-concept which relates to academic self-concept and to more subject-specific self-concept such as self-concept in mathematics or language. Self-concept in mathematics has been shown to predict various

measures of mathematic achievement while weak or even negative relation to achievement in language art (Marsh et al. 2006). Besides, the multifaceted structure involves academic self-concept and non-academic self-concept. The non-academic self-concept can be divided into social, emotional, and physical self-concept. In research, general self-concept often correlates with achievement with about .30, while academic self-concept tends to correlate higher with achievement, about .60 (Marsh 1992; Shavelson and Bolus 1982). Research supports a reciprocal effects model (REM) where prior academic achievement influences later self-concept and prior self-concept influences later achievement (Grygiel et al. 2016; Marsh and O'Mara 2007; Martin et al. 2010). Academic self-concept in specific school subjects has been shown to influence subsequent task choice; motivation; effort; persistence which, in turn; leads to improved achievement; and academic self-concept (Shavelson et al. 1976).

Students' success and failure in school may also depend on how well students manage in social situations such as relationship with peers, peer acceptance, and group membership (Wentzel 1991). Wentzel and Caldwell (1997) investigated the associations between group membership, peer acceptance, and peer relationship to academic achievement for sixth- and eighth-grade students. Of these, group membership was the best predictor of grades over time. Students' prosocial behavior, antisocial behavior, and emotional distress were found to be aspects that could explain the significant link between group membership and achievement.

### *The importance of students' motivation for achievement in school*

In self-determination theory, motivation is seen as the motive that regulates student's learning process and study behaviors, as well as the contexts that facilitates or hinders these regulations (Vansteenkiste et al. 2006). In achievement goal theory, mastery and performance goals are two main approaches of human aspiration (Ames 1984; Deci and Ryan 1985; Roberts 1992). However, Elliot and Harackiewicz (1996) demonstrated that the performance-goal approach is, in fact, two approaches: performance-approach and performance-avoidance. The avoidance construct is defined as students avoiding negative outcomes such as failure in schoolwork, being looked upon as stupid, or not understanding the material in courses (Derryberry and Rothbart 1997; Rothbart et al. 2000, 2001). If students have a motive disposition that is performance-approach or performance-avoidance such as strive to avoid failure or to attain success, it seems to be of less importance for achievement (Elliot and Harackiewicz 1996). Students holding an avoidance-approach may have a great motive to avoid failure and thus invest in the learning process and activities in school which may lead to success and higher achievements while affecting intrinsic motivation negatively (Deci and Ryan 1985). Students holding intrinsic goals or mastery goals in the learning process involve a deeper engagement in school activities, better conceptual learning, and higher persistence at learning activities, compared to students who hold extrinsic goals or performance-avoidance goals (Elliot and Harackiewicz 1996; Vansteenkiste et al. 2006). Thus, students' motives to improve in academic school subjects combined with motivational aspects of personality may be the core of the construct approach and avoidance (Elliot and Thrash 2010).

### **The importance of student background characteristics for achievement in school**

In many educational systems, girls receive higher grades when controlling for ability, compared to boys. When investigating final grades in compulsory school, Klapp and Cliffordson (2009) found that the gender differences in grades, benefitting girls, were almost fully

explained by girls showing a greater interest in learning. Hartley and Sutton (2013) found that girls and boys at an early age hold the belief that adults expect girls to do better in school, compared to boys. If these expectations and values are emphasized, the negative effect on boys' reading, writing, and achievement in math will increase, while girls' achievements are not affected by these kinds of beliefs.

The impact of socioeconomic status on students' achievement has been established by research and often explains about 10% of the variation in academic achievement on the individual level in schools in European countries (Yang 2003). Family socioeconomic status is mediated through the value and belief system within the family, the aspirations parents hold for their child, and their own academic efficacy. These aspirations and perceived efficacy for academic achievement have an indirect effect on the child's own self-concept and academic aspirations for the future.

However, recent research found that students, irrespective of socioeconomic status, were affected in a similar way by grading (Klapp et al. 2014) which supports previous research suggesting that being graded affects students differently primarily due to their cognitive ability and gender (Harlen and Deakin Crick 2002).

### **Theory of the relationship between summative assessments and achievement in school and personal resources**

The conservation of resources (COR) stress theory (Covington 2000; Frydenberg 2008; Hobfoll 1989, 2001) is an overall multidisciplinary theory which builds upon the beliefs that humans are trying to keep, develop, and gain personal resources in order to manage in difficult situations and throughout life. For students in school, resources are personal cognitive and non-cognitive skills such as academic and social self-concept, interest, and motivation. When students' resources are threatened, for example by failure in learning situations and in summative assessment situations and receiving bad school results, the loss of personal resources may cause emotional stress. The stress may, in turn, lead to negative and maladaptive learning strategies in order to avoid further losses and disparagement of schoolwork (Frydenberg and Lewis 2009). Students' self-concept is central in the COR stress theory for explaining consequences of failure in school. The COR stress theory adds to the social cognitive theory (Bandura 1986) by the focus on the overall consequences of riskful situations (for example, high-stake assessments) and stressful experiences (failures) in school.

In sum, the reviewed literature suggests that constructs such as academic and social self-concept and motivation may be of importance for understanding why low-ability students are negatively affected in their achievement when graded compared to low-ability students who are not graded. It also seems as if there are differences regarding gender and socioeconomic status and achievement.

### **The context of the study**

For 12 years, it was voluntary for Swedish municipalities to grade students in sixth grade (ages 12–13) or not (between years 1969 and 1981). In the subsequent grades (7–9), all students were graded. This unique circumstance combined with a large-scale national data project gathering information for cohorts of students made it possible to make comparisons between graded and ungraded students. The current study is based on Swedish data collected in the beginning of the 1980s; therefore, some aspects of the Swedish school system at that time will

be described. When data was collected, the Swedish school system was homogeneous and extensively centralized. There existed no free choice of school, but students belonged to a school due to their addresses and catchment area. Independent schools did not exist, and only a few private schools existed. There existed no tracking until the end of ninth grade (ages 15–16). Thus, a unique natural experiment existed in Sweden where some students received grades, while others did not.

During this period, students attended the same school from first to sixth grades. In seventh grade, they attended a new school, got new classmates, and were instructed by new teachers. They started to study biology, chemistry, and physics, which were new subjects. In grades 7–9, they were taught by new and different teachers in the different subjects. All students were graded at the end of seventh grade and, thereafter, at the end of every semester. Standardized national tests in Swedish, mathematics, and English were used to determine the level of achievement of the class in relation to the population on a national level, in order to support national comparability of the teacher-assigned grades. The grades were thus to be based on continuous classroom assessment in addition to the results on the standardized national tests.

External tests for the purposes of summative assessment may be of high stakes for students, teachers, and schools in many school systems. In Sweden, teacher-assigned grading is the primarily summative assessment used as an admission instrument for further studies in the educational system. Therefore, teacher-assigned grades hold a significant high-stake meaning to the Swedish students.

## **The design of the study**

The students had the same curriculum and studied the same subjects throughout school. About 50% of them were graded and 50% were not graded in sixth grade. All students took a questionnaire in sixth grade. The students had received instruction in the same subjects (from first or third grade), and they continued to receive instruction in seventh grade. By adding large-scale self-reported questionnaire data, it is possible to investigate if graded low-ability students received lower subsequent grades due to their lower self-concept and need to improve in academic subjects, compared to low-ability students who were ungraded in sixth grade.

## **Purposes**

The purpose of the study is to investigate if academic and social self-concept and motivation to improve in school mediate the negative effect of summative assessment (grades) on low-ability students' subsequent achievement in school. Differences due to personal background characteristics (cognitive ability, gender, and socioeconomic status) will be controlled for.

## **Method**

### **Subjects**

Data used in this study was retrieved from the ETF, a Swedish longitudinal project, which contains register and survey data for 10% nationally representative samples of individuals born between 1948 and 2004, compiled by Statistics Sweden (see Hämquist 2000). The purpose of

the ETF project is to gather data in order to conduct national evaluations of the school system and to be an access for research. The participants in the study were thus a 10% nationally representative sample of 8558 students born in 1967, the third cohort participating in the ETF project. The sampling was conducted by Statistics Sweden and was a stratified sampling procedure in two steps. First, municipalities were selected, and second, classes were selected. In total, 430 classes in 29 municipalities participated in the third cohort of the ETF project. The data used are subject grades in the seventh grade, when the participating students were 13 to 14 years old. Information on results on cognitive tests, gender, SES, and questionnaire data from the sixth grade was used.

## Measures

### *Not graded/graded*

During a period of 12 years, 1969 to 1981, the Swedish educational system let municipalities decide themselves if they wanted to grade their students in the sixth grade or not. Grading before seventh grade was abolished in 1982. Differences due to characteristics of the municipalities have been controlled for in previous studies (Klapp et al. 2014; Klapp 2015; Sjögren 2010).

When data was collected, teachers graded their own groups of students on a grading scale from 1 to 5, 1 being the lowest grade. The norm-referenced grades were calibrated to the standardized national test in Swedish, English, and mathematics and linked to the curve of normal distribution, at the population level. The point of reference was the mean performance of all students on a national level (subject and year). The norm-referenced grading system was constructed on the assumption that the distribution of grades was related to and followed a normal distribution pattern and that the grades for a class in the core subjects Swedish, English, and mathematics were to be based on students' results on centrally constructed standardized tests with a mean of 3 and a standard deviation of 1. However, the teacher-assigned subject grades for individual students were allowed to depart from the result from the standardized tests.

In all, 48.5% of the students were graded in the sixth grade ( $N = 4151$ ), whereas 51.5% were not graded in the sixth grade ( $N = 4407$ ). However, all the subjects received grades in the seventh grade.

Students who did not receive grades in the sixth grade were coded as 0 and students who did receive grades in the sixth grade as 1. Graded is a dummy variable.

### *Cognitive tests*

A cognitive test was used as a measure of students' cognitive ability. In the spring semester of sixth grade, all students in the study conducted three cognitive tests: one verbal, one spatial, and one inductive. The tests included opposite words, metal folding, and number sequences, respectively, each consisting of 40 tasks. The three subtests' correlations ranged from .41 to .51. These three tasks were summed into one total score, and this variable was standardized into z-scores (cognitive ability). The scale of cognitive ability ranged from  $-3.49$  to  $2.64$ . This kind of cognitive test has been shown to be a reliable measure of students' cognitive ability where differences due to gender and socioeconomic status are generally very small (Svensson 1971).

### *Socioeconomic status and gender*

SES is an index from parents' educational level, income, and occupation. SES is a continuous equidistant variable, with three categories. The SES variables is treated as a metric scale and

coded from 0 to 2 where 0 = a low socioeconomic status (SES III) ( $N = 3257$ ), 1 = medium socioeconomic status (SES II) ( $N = 3910$ ), and 2 = a high socioeconomic status (SES I) ( $N = 978$ ).

A dummy variable was constructed for gender (gender) where boys are coded as 0 ( $N = 4327$ ) and girls as 1 ( $N = 4231$ ).

### *Grade point average in seventh grade*

In seventh grade, all students were graded in 14 subjects. At the beginning of seventh grade, the students received instruction in the three new subjects: biology, chemistry, and physics. In previous studies (Klapp et al. 2014; Klapp 2015), the result showed that there were only small and non-significant effects of grading in these three new subjects; hence, they are not included in the GPA in the current study. The GPA is based on 11 grades. These are Swedish, English, mathematics, history, social science, geography, religion, music, craft, drawing, and athletics (GPA7).

### *Questionnaire data*

Students were given a questionnaire during the spring 1979/1980 in the sixth grade. Out of a large number of items, 14 items were considered meaningful in measuring certain students' academic and social self-concept and motivation to improve. These items have been used in the current study as indicators to create four factors (Table 1). Two academic self-concept factors were created: self-concept in mathematics (ScMa) and self-concept in Swedish (ScSw).

**Table 1** Students' self-concept in mathematics (ScMa) and Swedish (ScSw), self-concept in social situations (ScSocial), and students' motivation to improve in school subjects (MotImp) measured by questionnaire data

Items	Answers	
Self-concept in mathematics (ScMa)		
1. Do you think you are good at sums?	No–yes (2-point scale)	
2. Do you think that your teacher thinks that you are good at sums?		
Self-concept in Swedish (ScSw)		
3. Do you think that you are good at spelling?	No–yes (2-point scale)	
4. Do you think that your parents think that you are good at spelling?		
Self-concept in social contexts and social responsibility (ScSocial)		
5. If you had to arrange a party for your class. How good at it do you think you should be?	Not at all good–excellent (9-point scale)	
6. If you were elected onto the pupils council. How well could you cope with that?		
7. If you were asked to arrange and lead a play activity at school. How well could you cope with that?		
8. If you had to take the lesson when teacher was ill! How well could you cope with that?		
9. If your PE teacher wanted you to demonstrate a gym exercise. How well could you do that?		
10. How would other pupils evaluate your proposal for an Africa project?		
Motivation to improve (MotImp)		
11. Do you want to be better in school?		No–yes (2-point scale)
12. Do you often think that you would like to be better at reading?		
13. Do you often think that you would like to be better at doing sums?		
14. Do you often think that you would like to spell better?		

They were hypothesized to reflect students' academic self-concept in mathematics and Swedish by two items, each reflecting students' self-perception in counting and spelling such as "Do you think you are good at counting?" and "Do you think you are good at spelling?" The third factor, self-concept in social contexts (ScSocial) is hypothesized to reflect students' self-confidence in social contexts and social responsibility with indicators such as "If you had to take the lesson when teacher was ill. How well could you cope with that?" and "If you had to arrange a party for your class. How good at it do you think you would be?" The fourth factor (MotImp) reflects if students have a motive to improve in academic school subjects with indicators such as "Do you want to be better in school?" and "Do you often think that you would like to be better at doing sums?" The academic self-concept (ScMa and ScSw) and motivation to improve (MotImp) factors have binary response categories which may be seen as a weakness in the analyses, but by using modern powerful analytic procedures, it is possible to use the information in the manifest variables to construct factors (Muthén et al. 1997).

## Method of analysis

Confirmatory factor analysis (CFA) and structural equation modeling (SEM) were used in order to investigate the importance of academic and social self-concept and motivation to improve in academic school subject for the differential effect of grading. In a previous study (Klapp et al. 2014), multiple multivariate regressions were estimated to disentangle main effects as well as first-, second-, and third-order interaction effects of being graded on later achievement with control for background characteristics cognitive ability, gender, and socio-economic status. Only three of the first-, second-, and third-order cross product terms were significant (graded  $\times$  cognitive ability (INT 1), graded  $\times$  gender (INT 2), cognitive ability  $\times$  gender (INT 3)). The result showed no significant main effects of grading on students' later achievement. However, important interaction effects were found. Students with low scores on the cognitive test (low-ability students) were negatively affected by being graded in sixth grade in their subsequent grades in seventh grade, compared to low-ability students who were not graded. Effect sizes (Cohen's  $d$ ) were  $d = .30$  between ungraded and graded low-ability students and  $d = .14$  between ungraded and graded high-ability students (Klapp et al. 2014).

The variables have low intra-class correlations (ICCs) ranging from .003 to .045, except for cognitive ability (ICC = .078) and SES (ICC = .109). An ICC less than 5% is considered to be small and therefore not necessary to conduct a full multilevel analysis (Hox et al. 2010). Besides, whether or not to conduct a full multilevel analysis is due to the aim of the study rather than the characteristics of the data. Due to the aim of the current study, the complex option is considered an appropriate choice. To take account of effects of possible clustering of students in schools (school level), the "Complex" option offered by the Mplus program was used. This method compensates for disturbance in the chi-square and standard errors due to clustering effects, but it does not affect the estimates (Muthén and Muthén 1998–2012; Muthén and Satorra 1995). Because data is not available at the class level, ICC for class level was not possible to estimate. 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 and Muthén 1998–2012).

The first step involved the estimation of a measurement model (model A in Table 3) with the four factors (ScMa, ScSw, ScSocial, and MotImp) related to all of their respective indicators, with covariance between the factors. Second, a covariance model was estimated

(model B in Table 4) with covariance between all the variables. Next, a baseline model was estimated (model C1 in Table 5) with the main variables graded, cognitive ability, gender, and SES on GPA in seventh grade (GPA7). Then, a model was estimated with the main variables graded, cognitive ability, gender, and SES and the significant cross product terms graded  $\times$  cognitive ability (INT 1), graded  $\times$  gender (INT 2), and gender  $\times$  cognitive ability (INT 3) on GPA7 (model C2 in Table 5). Then, a SEM model with all the background variables (graded, cognitive ability, gender, and SES), the cross product terms (INT 1, INT 2, and INT 3), and the four factors (ScMa, ScSw, ScSocial, and MotImp) was estimated (model C3 in Table 5).

In the last modeling step, four SEM models (models D1 to D4 in Table 6) were estimated in order to examine the direct and indirect relations between academic and social self-concept and motivation to improve in academic school subjects and later achievement for graded and ungraded students (Fig. 1). The four factors (ScMa, ScSw, ScSocial, and MotImp) were included in the model each one at a time; hence, the model was estimated four times. The models were estimated with covariance between the independent variables. Tests of curvilinearity and homoscedasticity of the residuals were made, and no major deviations were found. The analyses were conducted with the Mplus program, version 5 (Muthén and Muthén 1998–2012).

The chi-square goodness-of-fit test and the root mean square error of approximation (RMSEA) were used as measures of model fit. The RMSEA takes both the number of observations and the number of free parameters into account and is strongly recommended as a tool when evaluating model fit (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 comparative fit index (CFI) measure was also used. This index should be as close to 1.0 as possible, and values below .95 are hesitant to accept (Bentler 1990). The Tucker-Lewis index (TLI) was used which is a measure similar to the CFI but has a penalty for models with many parameters. The TLI should be as close to 1 as possible, but values above .90 are considered acceptable (see, for example, Hu and Bentler 1995).

When creating factors, it is optimal for the indicators to have scales with several points. However, some of the indicators in the current study have 2-point scales but, by using the weighted least square mean and variance (WLSMV) parameter estimator in Mplus, it is possible to use 2-point scales (Muthén and Muthén 1998–2012; Brown 2006). The WLSMV parameter estimator uses a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistics that use a full weight matrix.

Missing information was handled using missing data modeling (Muthén et al. 1987), which is a method that makes the assumption that the data are *missing at random* (MAR), which infers that the procedure gives unbiased estimates when the missing is random is given the information in the data. This is a considerably less restrictive assumption than the assumption that the data are *missing completely at random*. High interrelations between the observed variables provide good possibilities to meet the MAR assumption (Schafer and Graham 2002).

## Results

In Table 2, the means and standard deviations for GPA from the seventh grade and cognitive ability, divided on not graded and graded students, boys and girls, and questionnaire data in the sixth grade, are presented.

**Table 2** Descriptive statistics for GPA in seventh grade, cognitive ability, and questionnaire data for not graded and graded students as well as for boys and girls, separately

Variable	Not graded = 0				Total		Graded = 1					
	Boys		Girls		N	%	Boys		Girls		Total	
	N	%	N	%			N	%	N	%	N	%
	2183	25.5	2224	26.0	4407	51.5	2144	25.0	2007	23.5	4151	48.5
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
GPA7	3.07	0.61	3.29	0.59	3.18	0.61	3.02	0.67	3.35	0.62	3.18	0.66
Cognitive ability <sup>a</sup>	-0.03	1.00	-0.03	0.99	-0.03	1.00	0.00	1.03	0.07	0.97	0.03	1.00
					Missing %							
SES I	978											
SES II	3919											
SES III	3257											
Questionnaire data												
14 items	7392–7892				7.8–13.6							

<sup>a</sup> The sum of the three cognitive tests standardized into z-scores (from -3.49 to 2.64)

There were missing observations for most variables. For the GPA7, the proportion of missing data was 7.9%. For SES and cognitive ability, the missing proportion was 4.8 and 10.5%, respectively. For the questionnaire data, the proportion of missing data ranged from 7.8 to 13.6%.

In order to examine if the advantage of the graded group is due to missing data, missing data must be taken into account in the analyses.

Independent *t* tests were computed in Mplus (Muthén and Muthén 1998–2012) in order to investigate possible effects of graded and gender (group differences) on cognitive ability, SES, and GPA7. The result showed no significant differences with respect to cognitive ability:  $t(7657) = 1.01, p = .32$ , and  $t(7657) = -0.12, p = .91$ , for graded and gender, respectively, and with respect to SES:  $t(8143) = -0.98, p = .33$ , and  $t(8143) = 1.86, p = .06$ , for graded and gender, respectively. Graded was not significant for GPA7:  $t(7878) = -1.63, p = .10$ . However, gender was significant for GPA7:  $t(7, 878) = 10.23, p = .00$ , which was expected.

**Measurement model of academic and social self-concept and motivation (model A)**

The analysis began by assessing the internal consistency of the scale for the four factors. Results showed that Cronbach’s alphas were .76, .77, .78, and .73 for ScMa, ScSw, ScSocial, and MotImp, respectively.

A measurement model (model A) was estimated with the four factors: ScMa, ScSw, ScSocial, and MotImp related to all their respective indicators, with covariance between the factors. This model suggests that the indicators were reasonable (see Table 3). All the standardized factor loadings were significant on the .001 level. The goodness-of-fit indices for this model were acceptable ( $\chi^2(34, 7927) = 780.55$ ; CFI = .936; TLI = .962; RMSEA = .053).

**Covariance model (model B)**

The next step in the modeling process was to estimate a covariance model with covariances between student background characteristics: variables (graded, cognitive ability, gender, and SES), the factors (ScMa, ScSw, ScSocial, and MotImp), and the dependent variable (GPA7) (model B in Table 4). The goodness-of-fit indices were acceptable ( $\chi^2(54, 8558) = 1375.25$ ; CFI = .901; TLI = .936; RMSEA = .053). First, there are negative relations between MotImp and

**Table 3** Standardized factor loadings for the four-factor model (model A) with covariance between the factors

Indicators	Latent variables			
	ScMa	ScSw	ScSocial	MotImp
1. Sums	.89			
2. Sums teach	.93			
3. Spelling		.98		
4. Spelling par		.86		
5. Party			.62	
6. Council			.69	
7. Leading			.65	
8. Lesson			.65	
9. Gym			.49	
10. Project			.61	
11. Better				.82
12. Read				.75
13. Sums				.80
14. Spell				.80
Goodness-of-fit for the four factor model	$\chi^2/df$ 780.55/34	CFI/TLI .936/.962	RMSEA .053	

All the standardized factor loadings are significant at the .001 level. For this model, the WLSMV parameter estimator, complex option, and missing data modeling have been used in Mplus

the other factors (ScMa, ScSw, and ScSocial), cognitive ability, gender, SES, and GPA7, which suggests that MotImp is a measure of low resource students (low ability, boys, low SES) with a motivation orientation that builds on the performance-approach as defined in the achievement goal theory (Elliot and Thrash 2010; Elliot and Harackiewicz 1996). There is a positive and significant relation between MotImp and graded, which suggests that students, who say that they want to improve their academic skills (performance-approach), receive higher subsequent grades if being graded in sixth grade. On the contrary, there is a negative relation between students who believe they are good in their academic skills (ScMa, ScSw) and being graded in sixth grade (graded). The relations between graded and the background variables and GPA7 are all non-significant. Since there are significant relations between graded and the academic self-concept factors as between cognitive ability and all four factors (ScMa, ScSw, ScSocial, and MotImp),

**Table 4** Standardized covariances for relations between student academic (ScMa, ScSw) and social self-concept (ScSocial), motivation for improvement (MotImp), background characteristics, and GPA in seventh grade (model B)

Variables	ScMa	ScSw	ScSocial	MotImp	Graded	Cognitive ability	Gender	SES	GPA7
ScMa	1	.230***	.432***	-.357***	-.052*	.430***	-.150***	.134***	.457***
ScSw		1	.212***	-.311***	-.047*	.143***	-.258***	.041**	.266***
ScSocial			1	-.171***	.036	.227***	-.079***	.138***	.299***
MotImp				1	.049*	-.388***	-.066***	-.149***	-.381***
Graded					1	.041	-.033	-.042	.003
Cognitive ability						1	.023	.252***	.632***
Gender							1	.026	.269***
SES								1	.287***
GPA7									1

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

there is reason to believe that there may exist interaction effects between cognitive ability and some of the background characteristics and GPA7 for graded and ungraded students (graded).

**Multiple regression models (models C1 to C3)**

First, a baseline model with only the main effects of graded, cognitive ability, gender, and SES on GPA7 was estimated (model C1). The standardized coefficients and *p* values are presented in Table 5. No significant main effect for graded on GPA7 was found. However, substantial main effects of cognitive ability, gender, and SES are shown in the result. Students with high cognitive ability, girls, and students with high socioeconomic status (SES I) receive higher GPA in seventh grade, compared to students with low cognitive ability, boys, and low socioeconomic status (SES III).

Then, a model was estimated with graded, cognitive ability, gender, SES, and three significant cross product terms [graded × cognitive ability (INT 1), graded × gender (INT 2), and gender × cognitive ability (INT 3)] on GPA7 (model C2 in Table 5). When including the cross product terms in the model (model C2), graded became significant for GPA7, which suggests that there exist differences between subgroups of students due to their cognitive ability and gender. Thus, graded low-ability students received lower grades in seventh grade, compared to ungraded low-ability students.

In the next step, a saturated regression model was estimated with the four factors (ScMa, ScSw, ScSocial, and MotImp) added to the previous model (C2) and related to GPA7 (see model C3 in Table 5). The result showed that the previous significant relation between graded and GPA7 became non-significant when the four factors were included in the model. The

**Table 5** Standardized regression coefficients for three multiple regression models (C1 to C3) with GPA in seventh grade as the dependent variable, and covariances between the independent variables

Variables	GPA7					
	Model C1		Model C2		Model C3	
	<i>β</i>	<i>t</i> value/ <i>p</i> value	<i>β</i>	<i>t</i> value/ <i>p</i> value	<i>β</i>	<i>t</i> value/ <i>p</i> value
Graded <sup>a</sup>	-.014	-1.10/.272	-.043**	-2.85/.004	-.014	-.74/.459
Cognitive ability <sup>b</sup>	.597***	72.29/.00	.564***	32.49/.00	.463***	21.15/.00
Gender <sup>c</sup>	.202***	20.15/.00	.170***	13.10/.00	.194***	14.63/.00
SES <sup>d</sup>	.134***	13.95/.00	.139***	14.69/.00	.118***	12.65/.00
INT <sup>e</sup> 1			.046**	3.15/.002	.020	1.13/.261
INT 2			.055**	3.10/.002	.036	1.42/.152
INT 3			-.016	-1.22/.221	-.027	-1.50/.134
ScMa					.192***	11.35/.00
ScSw					.069*	5.55/.00
ScSocial					.081***	6.53/.00
MotImp					-.072***	-5.47/.00
<i>R</i> <sup>2</sup>	.459***		.450***		.516***	

\**p* < .05; \*\**p* < .01; \*\*\**p* < .001

<sup>a</sup> Not graded = 0 and graded = 1

<sup>b</sup> The sum of the three cognitive tests, standardized into *z*-scores

<sup>c</sup> Boys = 0 and girls = 1

<sup>d</sup> A continuous equidistant variable coded as SES III = 0, SES II = 1, and SES I = 2

<sup>e</sup> INT 1 = graded × cognitive ability; INT 2 = graded × gender; INT 3 = gender × cognitive ability

goodness-of-fit was acceptable ( $\chi^2(60, 8558) = 1396.36$ ; CFI = .896; TLI = .934; RMSEA = .051). The CFI may be considered somewhat low which may be due to complexity of the model and the large number of parameters (Hu and Bentler 1999). The covariances between the factors ranged from  $-.17$  to  $.43$ . The strongest estimate among the factors for GPA7 was ScMa ( $\beta = .192$ ). ScSocial had the second strongest estimate to GPA7 ( $\beta = .081$ ), which suggests that students who believe they manage in social contexts and believe they can take social responsibility receive higher grades in seventh grade. MotImp had a significant negative estimate to GPA7 ( $\beta = -.072$ ), which shows that students who say they want to improve in academic school subjects receive lower GPA in seventh grade. The ScSw factor had the weakest estimate to GPA7 ( $\beta = .069$ ), however significant. The estimate of MotImp on GPA7 was negative, which suggests that students who say that they want to increase their academic skills receive lower subsequent grades (GPA7). This suggests that the MotImp factor is primarily a measure of a performance-avoidance orientation (Elliot and Harackiewicz 1996).

The estimates of the background characteristics (cognitive ability, gender, and SES) were all significant but somewhat lower compared to model C1. This result is reasonable since the four factors (ScMa, ScSw, ScSocial, and MotImp) explain variance in GPA7. All the cross product terms became non-significant when the factors were included in the model (C3).

### **Direct and indirect relations between student academic and social self-concept and motivation and GPA in seventh grade (models D1 to D4)**

Four models with direct and indirect relations between graded, GPA7, and the four factors (ScMa, ScSw, ScSocial, and MotImp), one at a time, were estimated. All the background characteristics (cognitive ability, gender, and SES) and the cross product terms (INT 1–3) were included. The goodness-of-fit indices were good for all the models. The standardized factor loadings and goodness-of-fit indices are presented in Table 6.

Here, an example of the D1 model (see Fig. 1) with direct and indirect relations is presented, so as to illustrate the models presented in Table 6. The relation from the ScMa factor to GPA7 is significant and rather substantial ( $\beta = .259$ ) while the direct relation of graded on GPA7 has become non-significant ( $\beta = -.017$ ). The indirect relation from graded to ScMa is significant and negative ( $\beta = -.068$ ) and is stronger compared to the direct relation from graded to GPA7 ( $\beta = -.043$ ) in the baseline model (model C2 in Table 5). This result indicates that self-concept in mathematics mediates the negative grading effect for low-ability students and that graded low-ability students had lower self-concept in mathematics, compared to ungraded low-ability students.

Similar result was found for the models with ScSw and MotImp (models D2 and D4) where the direct negative effect of grading on GPA7 became non-significant when the academic self-concept in Swedish and the motivation to improve academic skills were taken into account. The negative effect of grading in sixth grade for low-ability students seems to be due to their lower self-concept in mathematics (ScMa) and in Swedish (ScSw) and their stronger motive to improve their academic skills (MotImp). Social self-concept (ScSocial) did not mediate the negative grading effect for low-ability students.

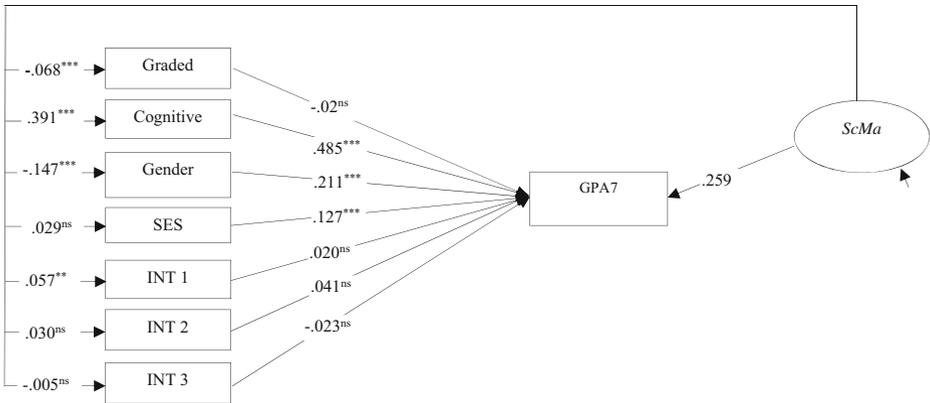
In sum, when taking into account students' self-concept in mathematics and Swedish and the motive to improve academic skills, the negative grading effect on subsequent achievement became non-significant, when controlling for the background characteristics. This result suggests that students' self-concept in mathematics and Swedish and motivation to improve

**Table 6** Relations for the models D1 to D4 with direct and indirect relations between graded and the four factors (ScMa, ScSw, ScSocial, and MotImp) on GPA in seventh grade, and covariances between the independent variables

Model	Latent variable	Graded	Cognitive ability	Gender	SES	INT 1	INT 2	INT 3	GPA7	$\chi^2/df$	RMSEA/CF/TLI
D1	ScMa Graded	-.068***	.391***	-.147***	.029	.057**	.030	-.005	.259***	22.88	.020
D2	ScSw Graded	-.054*	.133***	.189***	-.001	.029	.028	-.016	-.017	5	.997/.993
D3	ScSocial Graded	.006	.196***	-.085***	.091***	.004	.035	.005	.124***	21.32	.022
D4	MotImp Graded	.050*	-.320***	-.045*	-.052***	-.028	-.002	-.051**	-.028	4	.997/.991
									.175***	600.55	.036
									-.045*	49	.957/.938
									-.142***	427.67	.055
									-.029	16	.945/.945

INT 1 = graded × cognitive ability; INT 2 = graded × gender; INT 3 = gender × cognitive ability

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$



**Fig. 1** A structural equation model (model D1) with direct and indirect relations between the socioemotional factor ScMa and GPA7 for not graded and graded students. This is a demonstration of how the models in Table 6 can be interpreted

their academic skills in school mediate the negative grading effect for low-ability students on their subsequent grades.

## Discussion and conclusions

The main purpose of the current study was to examine if academic and social self-concept and motivation to improve academic skills explained the negative effect of summative assessment (grades) on low-ability students' subsequent achievement in school. The main contribution of the study is that the negative grading effect for low-ability students was explained by their lower self-concept and need for improvement, which may be due to them having experienced a stronger summative assessment regime which has affected their self-concept negatively (Covington 2000; Frydenberg 2008; Hobfoll 1989). The result suggests that experiencing a summative assessment regime and being a student with low cognitive ability lead to lower subsequent grades compared to experiencing a less summative assessment regime and being a low-cognitive ability student. The result shows that low-ability students who experience a summative assessment regime (grading), to a larger extent, believe they are bad in mathematics and Swedish and that they need to improve their academic skills and receive lower subsequent grades, compared to low-ability students who does not experience a summative assessment regime (grading). This result suggests that the negative grading effect for low-ability students seems to be explained by their lower academic self-confidence and beliefs that they need to improve their academic skills.

According to research, students' self-concept regulates students' learning and how they manage difficult subject matters (Bandura 1986; Wentzel and Caldwell 1997; Wentzel et al. 2010) which may be one explanation for the predictive power of students' self-concept in mathematics and Swedish after controlling for cognitive ability, gender, and socioeconomic status. The predictive power of students' self-concept in social situations on later achievement shown in this study does not clearly support the findings of Wentzel and Caldwell (1997) and Wentzel et al. (2010) that suggest that students' success in school is partly due to how well students manage in social situations and with peers. One possible explanation may be that the

current study controls for cognitive ability and takes into account differential effects of summative assessment on low-ability students' subsequent achievement.

### **The direct and indirect associations between grading, academic, and social self-concept and motivation to improve in school and achievement**

The negative grading effect for low-ability students became non-significant when students' self-concept in academic subjects and motive to improve their academic skills were taken into account. Students who were graded in sixth grade had probably experienced more feedback of summative character, which probably increased the risk of effects on students' self-concept and motivation. For low-ability students, the effects of a summative assessment regime seem to be more negative, whereas for high-ability students, the consequences of summative assessment regime seem to be neither negative nor positive (Klapp et al. 2014; Klapp 2015). Hence, students in this study have experienced different amounts of success and failure in school (graded and ungraded environments) which, in turn, seem to influence students' self-concept, motivation, emotions, and self-determination (Ryan and Deci 2000a, b). Reciprocal relations between assessment and non-cognitive competencies may explain the grading gap between graded low-ability students and ungraded low-ability students and why they are affected differently by grading (Marsh and O'Mara 2007; Martin et al. 2010; Pekrun et al. 2002). In line with several researchers (Purdie and Hattie 1996; Schunk 1996), it seems reasonable to believe that in classrooms where grading took place, there was a risk that teachers had a more rigid and differential structure regarding assessment practices which increases the risk of failure for low-ability students (Covington 2000; Hobfoll 1989, 2001). If this was the case, students may have developed a stronger focus on performance goals such as an avoidance orientation (Elliot and Harackiewicz 1996; Elliot and Thrash 2010) which may have affected their academic and social self-concept and learning.

Students who have experienced failures in school, both academic and social, may lose resources such as self-confidence in academic subjects and in social situations (Covington 2000; Frydenberg 2008; Hobfoll 1989, 2001). According to the COR stress theory, students need to keep, gain, and develop a sense of self-worth and positive self-confidence in order to believe that they can manage, learn, and achieve in school (Covington 2000; Hobfoll 1989, 2001). The fact that three of the non-cognitive competencies in this study mediated the negative effect of grading for low-ability students supports the COR stress theory that students' non-cognitive or affective competencies (resources) are important determinants for students' success in school.

### **Educational implications of the results**

Research has shown that children already at an early age have developed a self-image in line with expectancies and values their parents and teachers hold (Hartley and Sutton 2013; Molden and Dweck 2006). Therefore, it is of importance to create a learning situation in school which develops students' self-confidence in a positive way. Thus, schools need to address both cognitive and non-cognitive skills in order to raise students' learning and achievement (Duckworth and Seligman 2005; Durlak et al. 2011; Dweck and Yeager 2012). The result suggests that a summative assessment practice in school affects low-ability students negatively (ungraded low-ability students receive higher subsequent grades) and that it might not be the summative assessment such as a grade itself that affects low-ability students

negatively but the expectations adults have on students' capacity to learn, how teachers and parents act when students fail or succeed in school, how teachers talk about learning and assessments, and what adults believe is possible for certain children and students. It seems as one of the most important missions for teachers is to support students' success in school and to avoid an instruction and assessment practice that involves high risks of failures.

### **The generalizability of the findings**

The data in the current study was gathered in 1980, and at that time, Sweden had a norm-referenced grading system where 1 was the lowest and 5 the highest grade. Students who received the lowest grade could still continue to higher levels within the educational system. Today, Sweden has a criterion-referenced grading system with a fail step which makes it harder for students to continue to higher levels within the educational system. At least 8 or 12 passing grades are required to enter a vocational or theoretical program in upper secondary education. Thus, in today's criterion-referenced grading system, grades have stronger high-stake meaning to students, compared to grades in the previous norm-referenced grading system. If the study was replicated on data gathered today, it is reasonable to believe that the negative grading effect would be as strong as in the current study or even stronger.

Besides, the results found here (for a norm-referenced grading system) are in line with research results based on research conducted in different educational and grading systems and over time (Duckworth and Seligman 2005; Durlak et al. 2011; Dweck and Yeager 2012; Heckman and Kautz 2012; Molden and Dweck 2006). It seems as if the importance of students' non-cognitive skills for their later achievement is a robust finding, across time and over grading systems. Even though the quality of the data in the current study may be considered high due to the nationally representative sample and the alignment with previous research results, the result from the current study should be generalized with some caution and the study should be replicated in other educational settings.

### **Limitations and further research**

One limitation in the current study is that the data was gathered in the beginning of the 1980s when students were graded in a norm-referenced grading system. Even though it can be argued that the negative grading effect may exist in a criterion-referenced grading system, more research evidence is needed. Therefore, in an ongoing study, data from the criterion-referenced grading system is used to examine how grading in sixth grade affects students' later achievement in school. Between the years 1994 and 2010, grading did not exist until students reached the eighth grade (grading existed only in eighth and ninth grades), and Sweden introduced grades in sixth grade in 2011/2012; thus, students receive grades each semester between sixth and ninth grades. This circumstance makes it possible to compare ungraded and graded students' subsequent achievement for different cohorts: before and after the grading reform in 2011/2012.

Another limitation of the current study concerns the questionnaire. Since the questionnaire was developed already in the 1960s with some changes made in the beginning of the 1980s, it is reasonable to believe that other possible important non-cognitive competencies have been left out due to the restricted instrument. The items in the questionnaire were created in line with research at the time, and some of the items have a 2-point scale which may limit the degree of information. In more recent cohorts in the ETF project, items with 5-point Likert scales are

available as well as items measuring a wider range of non-cognitive competencies, for example negative emotions. Therefore, in future research, it would be of interest to investigate other possible non-cognitive competencies such as resilience, grit, and negative emotions in order to deepen the understanding of the negative grading effect on low-ability students' achievement.

Besides, further research should investigate the direct and indirect relations between non-cognitive competencies, summative assessments, subsequent achievement, and success in the working life and overall adult life.

**Acknowledgements** The work is a part of the COMP (Grades: comparability, prognostic validity and effects on learning) project which is financially supported by the Swedish Research Council.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alexander, R. (Ed.). (2010). *Children, their world, their education. Final report and recommendations of the Cambridge primary review*. London: Routledge.
- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: a cognitive motivational analysis. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 177–207). New York: Academic Press.
- Artes, J., & Rahona, M. (2013). Experimental evidence on the effect of grading incentives on student learning in Spain. *Journal of Economic Education*, 44(1), 32–46.
- Azmat, G., & Iriberry, N. (2009). *The importance of relative performance feedback information: evidence from a natural experiment using high school students. CEP discussion paper no 915, Centre for Economic Performance*. London: London School of Economics and Political Science.
- Bandiera, O., Larcinese, V., Rasul, I. (2008). *Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on performance*. Mimeo.
- Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs: Prentice Hall.
- Becker, W. E., & Rosen, S. (1992). The learning effect of assessment and evaluation in high school. *Econ Educ Rev*, 11(2), 107–118.
- Bentler, P.M. (1990). Comparative indexes in structural models. *Psychology Bulletin*, 107, 238–246.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1–14.
- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, C. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, 100(3), 525–534.
- Carrillo-de-la-Peña, M. T., Baillès, E., Caseras, X., Martínez, A., Ortet, G., & Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education: Theory and Practice*, 14(1), 61–67.
- Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & van der Vleuten, C. P. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education*, 15(5), 695–715.
- Conger, D., & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The ANNALS of the American Academy of Political and Social Science*, 627(1), 184–214. doi:10.1177/0002716209348751.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: an integrative review. *Annual Review of Psychology*, 51, 171–200.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.

- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.
- Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology*, 9, 633–652. doi:10.1017/S0954579497001375.
- Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939–944.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432.
- Dweck, C. S., & Leggett, E. (1988). Asocial-cognitive approach to motivation and personality. *Psychological Review*, 95, 256–273.
- Dweck, C. S., & Yeager, D. S. (2012). Mindsets that promote resilience: when students believe that personal characteristics can be developed. *Educational Psychologist*, 47(4), 302–314.
- Eccles, J. (1984). Sex differences in achievement patterns. In T. Sonderegger & R. A. Dientsbier (Eds.), *Nebraska symposium on motivation, 1984: psychology and gender* (Vol. 32, pp. 97–132). Lincoln: University of Nebraska Press.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: a mediational analysis. *Journal of Personality and Social Psychology*, 70, 461–475.
- Elliot, A. J., & Thrash, T. M. (2010). Approach and avoidance temperament as basic dimensions of personality. *Journal of Personality*, 78, 865–906.
- Frydenberg, E. (2008). *Adolescent coping: advances in theory, research and practice*. London & New York: Routledge, Taylor & Francis Group.
- Frydenberg, E., & Lewis, C. (2009). The relationship between problem-solving efficacy and coping among Australian adolescents. *British Journal of Guidance and Counselling*, 37(1), 51–64.
- Grygiel, P., Modzelewski, M., & Pisarek, J. (2016). Academic self-concept and achievement in polish primary schools: cross-lagged modelling and gender-specific effects. *European Journal of Psychology Education*. doi:10.1007/s10212-016-0300-2.
- Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1\*). In: *Research evidence in educational library*. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education
- Härmqvist, 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 behavioural science*. Forskningsrådsnämnden: Stockholm.
- Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development*, 84, 1716–1733.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19, 451–464.
- Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, 26(3), 276–295. doi:10.1016/j.appdev.2005.02.007.
- Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *American Psychologist*, 44(3), 513–524.
- Hobfoll, S. E. (2001). The influence of culture, community, and the nested-self in the stress process: advancing conservation of resources theory. *Applied Psychology: An International Review*, 50(3), 337–421.
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 76–99). Thousand Oaks: Sage.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, 30, 11–21. doi:10.1016/j.lindif.2013.12.003.
- 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: Sage.
- Klapp, A. (2015). Does grading affect Educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy and Practice*, 22(3), 302–323.
- Klapp, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*, 15(1), 1–23.

- Klapp, A., Cliffordson, C., & Gustafsson, J.-E. (2014). The effect of being graded on later achievement: evidence from 13-year olds in Swedish compulsory school. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 1-19. doi:10.1080/01443410.2014.933176.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84(1), 35–42.
- Marsh, H. W., & O'Mara, A. (2007). Reciprocal effects between academic self-concept, self-esteem, achievement and attainment over seven adolescent years: unidimensional and multidimensional prospects of self-concept. *Personality and Social Psychology Bulletin*, 34, 542–552.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: construct validation and relations to well-being and achievement. *Journal of Personality*, 74, 403–456.
- Martin, J. A., Colmar, H. S., Davey, A. L., & Marsh, W. H. (2010). Longitudinal modelling of academic buoyance and motivation: do the '5Cs' hold up over time? *British Journal of Educational Psychology*, 80, 473–496.
- Molden, D. C., & Dweck, C. S. (2006). Finding "meaning" in psychology. A lay theories approach to self-regulation, self-perception and social development. *American Psychologist*, 61(3), 192–203.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles: Muthén & Muthén.
- Muthén, B., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimate latent variable relations with a comprehensive measurement model. *Psychometrika*, 60(4), 489–503.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431–462.
- Muthén, B., du Toit, S.H.C. & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished technical report. Retrieved 2016-04-20 from <https://www.statmodel.com/wlscv.shtml>.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155–175.
- Nicholls, J. G., & Miller, A. T. (1984). Development and its discontents: The differentiation of the concept of ability. In J. G. Nicholls (Ed.), *Advances in motivation and achievement: vol. 3. The development of achievement motivation* (pp. 185–218). Greenwich: JAI Press.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: a program of quantitative and qualitative research. *Educational Psychologist*, 37, 91–106.
- Pollard, A., Triggs, P., Broadfoot, P., Mcness, E., & Osborn, M. (2000). *What pupils say: changing policy and practice in primary education (chapters 7 and 10)*. London: Continuum.
- Purdie, N., & Hattie, J. (1996). Cultural differences in the use of strategies for self-regulated learning. *American Educational Research Journal*, 33(4), 845–871.
- Roberts, G. C. (1992). Motivation in sport and exercise: conceptual constraints and convergence. In G. Roberts (Ed.), *Motivation in sports and exercise* (pp. 3–29). Champaign: Human Kinetics.
- Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: origins and outcomes. *Journal of Personality and Social Psychology*, 78, 122–135. doi:10.1037//0022-3514.78.1.122.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. *Child Development*, 72, 1394–1408. doi:10.1111/1467-8624.00355.
- Ryan, R. M., & Deci, E. L. (2000a). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist*, 55, 68–78.
- Ryan, R. M., & Deci, E. L. (2000b). When rewards compete with nature: the undermining of intrinsic motivation and self-regulation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: the search for optimal motivation and performance* (pp. 13–54). New York: Academic.
- Schafer, L. J., & Graham, W. J. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7(2), 147–177.
- Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, 33, 359–382.
- Shavelson, R. J., & Bolus, R. (1982). Self-concept: the interplay of theory and methods. *Journal of Educational Psychology*, 74(1), 3–17.
- Shavelson, R. J., Hubner, J. J., & Stanton, J. C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research*, 46, 407–441.

- Sjögren, A. (2010). *Graded children—evidence of long-run consequences of school grades from a nationwide reform*. Uppsala: IFAU – Institute for Labour Market Policy Evaluation.
- Svensson, A. (1971). *Relative achievement. School performance in relation to intelligence, sex and home environment. Göteborg studies in educational science, Vol. 6*. Göteborg: Acta Universitatis Gothoburgensis.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education, 31*(5), 505–521.
- Vansteenkiste, M., Lens, M., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: another look at the quality of academic motivation. *Educational Psychologist, 41*(1), 19–31.
- Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development, 62*, 1066–1078.
- Wentzel, K. R., & Caldwell, K. (1997). Friendships, peer acceptance, and group membership: relations to academic achievement in middle school. *Child Development, 68*(6), 1198–1209.
- Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology, 35*(3), 193–202. doi:10.1016/j.cedpsych.2010.03.002.
- Yang, Y. (2003). *Measuring socio-economic status and its effect at individual and collective levels: a cross-country comparison (Göteborg studies in educational science, vol. 193)*. Göteborg: Acta Universitatis Gothoburgensis.

**Alli Klapp**, Department of Education and Special Education, University of Gothenburg Box 300 SE 405 30 Gothenburg, Sweden. E-mail: alli.klapp@ped.gu.se. Phone work: (46) 46 738 558830. alli.klapp@ped.gu.se. [www.ips.gu.se](http://www.ips.gu.se)

*Current themes of research:*

The themes of my research are mainly within the assessment and evaluation field and for compulsory and upper secondary education. I conduct research concerning what grades measure in terms of students' cognitive skills and non-cognitive competencies and gender differences in grades, and how summative assessment such as grading and testing affects students' subsequent learning and achievement. Several studies focus on the predictive validity of students' non-cognitive competencies for achievement in school and for success in working life. I have recently cooperated with Professor Henry Levin and his team of researchers at the Columbia University in New York, concerning the effects of developing students' socioemotional competencies for achievement in school. In my research, I conduct large-scale studies using national representative samples ( $N = 10,000$ ) or population data (whole cohorts of students in Sweden,  $N = 100,000$ ) containing measures of students' cognitive ability, gender, and socioeconomic status, and grades, national test results, and self-reports from students, teachers, and parents. Confirmatory factor analysis (CFA) and structural equation modeling (SEM) are the main analyses techniques being used.

*Most relevant publications in the field of Psychology of Education:*

- 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. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation, 15*(1), 1–23.
- Klapp Lekholm, A. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research, 6*(55), 587–608.
- Klapp, A., Cliffordson, C., & Gustafsson, J-E. (2014). The effect of being graded on later achievement: evidence from 13-year olds in Swedish compulsory school. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 1–19. DOI: 10.1080/01443410.2014.933176.
- Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy and Practice, 22*(3), 302–323.
- Belfield, C., Bowden, B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis, 6*(3), 508–544.
- Klapp, A. (2016). The importance of self-regulation and negative emotions for predicting educational outcomes—evidence from 13-year olds in Swedish compulsory and upper secondary school. *Learning and Individual Differences, 52*, 29–38.