GOTHENBURG STUDIES IN EDUCATIONAL SCIENCES 356

Dimensionality and Predictive validity of school grades:

The relative influence of cognitive and social-behavioral aspects

Cecilia Thorsen

UNIVERSITY OF GOTHENBURG
ACTA UNIVERSITATIS GOTHOBURGENSIS

Dimensionality and predictive validity of school grades

GOTHENBURG STUDIES IN EDUCATIONAL SCIENCES 356

Dimensionality and predictive validity of school grades:

The relative influence of cognitive and social-behavioral aspects

Cecilia Thorsen

UNIVERSITY OF GOTHENBURG
ACTA UNIVERSITATIS GOTHOBURGENSIS

Thesis in Education at the Department of Education and Special Education

ISBN 978-91-7346-797-1
ISBN 978-91-7346-798-8 (pdf)
ISSN 0436-1121

The thesis is also available in full text on:
http://hdl.handle.net/2077/36586

Distribution:
Acta Universitatis Gothoburgensis
Box 222
405 30 Göteborg
acta@ub.gu.se

Photographer cover: Frida Thorsen

Print:
Ale Tryckteam, Bohus, 2014

Abstract

Title:	Dimensionality and predictive validity of school grades: The relative influence of cognitive and social-behavioral aspects
Author:	Cecilia Thorsen
Language:	English with a Swedish summary
ISBN:	978-91-7346-797-1 (tryckt)
ISBN:	978-91-7346-798-8 (pdf)
ISSN:	0436-1121
Keywords:	predictive validity, dimensionality, criterion-referenced grades, norm-referenced grades, social-behavioral aspects, crystallized abilities, fluid abilities

The purpose of the thesis is to investigate the relative influence of cognitive and social-behavioral aspects on compulsory school grades and the importance of the different dimensions for the predictive validity of grades. Data is retrieved from the Gothenburg Educational Longitudinal Database (GOLD) and the Evaluation Through Follow-up (ETF) database. The sample in Study I consisted of three cohorts each of about 100 000 students in Grade 9, in Study II of about 4000 students in Grade 9, and in Study III of about 9000 students who were followed-up through compulsory school. All analyses were conducted using structural equation modelling (SEM).

Both criterion-referenced and norm-referenced compulsory school grades were found to be multidimensional, reflecting both subject-specific dimensions and a common-grade dimension, cutting across grades and teachers. The common-grade dimension, which in previous research has been found to be related to social-behavioral aspects, contributed to predict study success in upper secondary school, indicating that social-behavioral aspects partly contribute to explain the predictive power of school grades.

The influence of cognitive aspects was substantial. Fluid abilities had a continuous direct influence on the development of knowledge and skills throughout compulsory school, which is in line with the predictions from Cattell's (1987) Investment theory. Substantial indirect effects of fluid abilities on school grades were found, although no direct effects. In sum the results in the present thesis show that both cognitive and social-behavioral aspects contribute to explain the predictive validity of school grades.

Table of contents

Acknowledgments

Pursuing PhD studies has been a journey of a lifetime on which there have been both steep hills as well as beautiful landscapes. I am very grateful to the many people who have guided me on this journey.

First and foremost, I would like to express my gratitude to my main-supervisor Christina Cliffordson and my co-supervisor Jan-Eric Gustafsson, without whom I would sometimes have been stumbling blindly. Christina, I am endlessly grateful to you for keeping me on the right track and for always being available for answering questions and for giving invaluable advice. Jan-Eric, I would like to give my sincere thanks to you for introducing me to the world of cognitive abilities and the world of statistics. I would also like to acknowledge my gratitude to Alli Klapp who was my co-supervisor in the early stages. Thank you for being such a great support and for reading and commenting on my manuscripts.

I would also like to offer my sincere thanks to the members of the FUR-group for welcoming me to your ranks and for always providing support and sharing your knowledge. A special thanks to Alastair Henry for being a great friend and colleague and not only for scrutinizing my language, but also giving invaluable advice. I would also like to thank Stefan Johansson who has been a great friend and colleague and who generously has shared his knowledge.

Furthermore, I would like to thank Gudrun Erickson who was not only the discussant at my planning seminar, but at several stages also discussed the findings with me. I would also like to acknowledge my gratitude to Kristian Ramstedt, discussant at the mid-stage seminar, and Martin Bäckström, discussant at my final seminar.

Finally, I am grateful to my friends and family. A special thank you to my dear friend Åsa Windfäll for providing advice on life as a doctoral student even though our areas are quite different. My deepest gratitude and love goes to Morten, Frida and Oscar, you have always been by my side and without you I would not have finished this journey.

Hunnebostrand, August, 2014

1 Introduction

In Sweden grades are used as instruments for selection to the next level in the educational system, both to upper secondary school and to higher education studies. When grades are used for selection purposes they are supposed to be both reliable and valid. This includes aspects such as being fair and comparable between teachers, between schools and over time. It is also important that an instrument used for selection is a reliable indicator that the students who are selected have the proper prerequisites in terms of knowledge and skills, and that the instrument is able to select the students who are best equipped to handle the education. However, solutions used for rank-ordering students for selection to the next level in the educational system differ, among other things, with respect to whether capacity for studies or previous achievement should constitute the basis for selection. According to Lohman's (2004) distinctions different instruments which are used for selection can be placed on a fluid-crystallized ability continuum. Instruments assessing cognitive abilities or capacity for studies, such as the SAT and SweSAT, can be placed on the fluid end of such a continuum, while on the crystallized end can be located tests measuring declarative knowledge and the ability to solve familiar problems, such as achievement tests. A large amount of research has investigated the advantages and disadvantages of different solutions for selection of students.

Traditionally tests measuring different dimensions of cognitive abilities or capacity for studies have been considered as more reliable measures of future educational success than measures of achievement (Geiser & Santelices, 2007). However, when considering prediction and predictive power, measures of achievement (in particular school grades from the previous educational level) typically display higher predictive power than tests measuring capacity for schooling, such as the SAT and SweSAT (e.g. Bowen, Chingos & McPherson, 2009; Geiser & Santelices, 2007; Gustafsson, 2003; Gustafsson & Carlstedt, 2006). Considering research on prediction from a life-span perspective, the pattern of predictive validity seems to indicate that fluid abilities have a great impact on the development of knowledge and skills in the early school years, much in line with Cattell's (1971, 1987) Investment theory, but when students

go through the school system they increasingly seem to rely on crystallized abilities, which could explain why measures of declarative knowledge and achievement such as grades have stronger predictive power than measures of fluid abilities.

There is also research which emphasizes the importance of personality factors for the development of knowledge and skills in adult samples (e.g. Ackerman, 1996). One explanation for the typically strong predictive validity of school grades could be that they reflect both fluid and crystallized abilities, as well as a broader array of knowledge and skills, as well as factors which relate to personality (Gustafsson & Carlstedt, 2006). The breadth of knowledge and skills represented by grades seems to play a part in explaining their predictive power, as well as their relation to social-behavioral aspects (Almlund, Duckworth, Heckman & Kautz's, 2011).

As indicated by several studies, grades are by no means unproblematic measures. Rather, they seem to be multidimensional measures reflecting both cognitive and social-behavioral aspects (e.g. Alexander, 1935; Gustafsson & Balke, 1993; Klapp Lekholm & Cliffordson, 2008; 2009). Even though grades are of great importance for students' opportunities, underscoring the importance of fairness and comparability, there are indications that these principles can be challenged. Research on teachers' grading practices indicates that grades may suffer both from construct underrepresentation and construct irrelevant variance (e.g. Brookhart, 1991; 1993). In addition, grades have been shown to suffer from grade inflation (Cliffordson, 2004a).

Nevertheless, measures of achievement, and in particular school grades, have stronger predictive validity than measures of academic promise (Atkinson, 2001; Cliffordson, 2008; Geiser & Santelices, 2007; Gustafsson & Carlstedt, 2006). Measures of achievement are also more closely aligned to curricular content and they signal to students that it is beneficial to put effort into schoolwork and what is important to learn in school. Such measures have also been found to be fairer in being less connected to socioeconomic background, and it could be argued that measures of achievement are better incentive devices for both students and schools (Atkinson, 2001).

Research has also shown that predictive power pertains both to norm- and criterion-referenced grades. Grades from both grading systems were more powerful in predicting study success than tests measuring capacity for studies (Cliffordson, 2008). However, the reason why grades have better predictive validity than tests measuring different dimensions of cognitive abilities and

tests measuring achievement has not been fully clarified. One possible explanation for this pattern of predictive validity could be, as discussed above, that grades are multidimensional measures reflecting different dimensions of abilities, encapsulating both fluid and crystallized abilities, as well as a broader array of knowledge and skills and, in addition, social-behavioral aspects. (Bowers, 2011; Cliffordson, 2008; Gustafsson, 2003; Gustafsson & Carlstedt, 2006; Klapp Lekholm & Cliffordson, 2008, 2009). It is possible that both cognitive and social behavioral aspects contribute to explain the predictive power of school grades.

The present study investigates the influence of cognitive, social-behavioral aspects of school grades and the importance of the different dimensions for their predictive validity. The thesis consists of three empirical studies and a theoretical framework. First a contextual background is given, followed by presentations of previous research and theoretical premises and foundations. These are followed by a summary of the three empirical studies, a discussion of the findings and a conclusion.

Aims

The primary aim of the present thesis is to better understand the predictive power of compulsory school grades. The research questions are based on the body of research showing that measures of achievement, and in particular school grades, are typically better predictors of future achievement than measures intended to measure different cognitive dimensions. It is a reasonable assumption that this predictive pattern could, in part, be explained by the multidimensional nature of school grades, that is to say the encapsulating of both cognitive and social-behavioral aspects. Consequently, the aim is to investigate the influence of cognitive and social-behavioral aspects in school grades and the relative importance of these aspects for their predictive validity. Another aim is to investigate the stability of the dimensionality of grades, as well as the potential differences due to gender and educational background.

A conceptual framework discussing the dimensionality and predictive validity of school grades, as well as studies on teachers' grading practices, is used as a backdrop for the analyses on the influence and importance of both knowledge and skills and social-behavioral aspects. Cattell's (1987) Investment Theory, Gustafsson and Carlstedt's (2006) Encapsulation Theory and

Ackerman's PPIK Theory (1996) are used as the theoretical basis in order to investigate the influence of cognitive aspects on school grades. In particular, Cattell's (1987) Investment theory is tested in order to explain individual differences in acquisition of knowledge and skills as a result of the investment of cognitive resources. Investment theory, Encapsulation theory and the PPIK theory are also used as the basis for the interpretation of both cognitive and social-behavioral aspects in grades and their relative importance for the predictive validity.

Issues of validity in relation to the function of grades and teachers' grading practices are also considered. In particular the validity of grades is discussed in relation to Messick's (1989) validity framework. Detailed aims are given in relation to the presentation of each study included in the present thesis

2 Contextual background

Definitions

Measures of achievement are often used to measure educational outcomes. These are constituted by, for example, grades and results on national tests. Statistics on educational results measured by grades and national tests are, in Sweden, typically published on an annual basis by the National Agency for Education (e.g. National Agency for Education, 2013). Grades and national tests are also used in Sweden to evaluate the quality of schooling. Other types of external tests, such as the Progress in International Reading Literacy Study (PIRLS) and Program of International Achievement (PISA), have also proven to be of great importance for how the results of schools are evaluated.

Achievement is an important goal of schooling, but there are also overarching societal goals which are considered important and which are often of a social and behavioral nature. These overarching social and behavioral goals are emphasized in the curriculum as important aspects of schooling. Social and behavioral aspects have, just as measures of cognitive ability and achievement, been proven to be good predictors of future academic performance (e.g. Rosander, 2012). Still, even though social and behavioral aspects have been demonstrated to be important in many different ways, as well as predicting future achievement, they do not form a part of either grades or national tests.

Not only does the underlying structure, definition and measurement of such social and behavioral aspects involve complex issues, but the terminologies used to refer to such aspects differ between different researchers. Sometimes they are referred to as 'non-cognitive' or 'social and behavioral' aspects, and sometimes as 'affective' or 'socio-emotional' aspects (Levin, 2011). These different definitions can be grounded in differing underlying structures and might thus have different meanings in different contexts. However, the main purpose of this thesis is to distinguish these aspects from the knowledge and skills tested in knowledge tests and which are a part of the criteria in grades. The term 'social and behavioral aspects' is used within this thesis since it is a better representation of aspects which could be valued in school. Social aspects could be aspects such as being able to

cooperate with others, being able to take turns and being able to adopt other peoples' perspectives. While behavioral aspects could be aspects such as attendance, demonstrating effort and engagement, and being able to organize school work. The concept 'non-cognitive' is a more undefined concept, which is why 'social and behavioral' is preferred. Social and behavioral aspects encapsulate aspects which are important for daily work in the classroom and also aspects which are emphasized in the curriculum and which could be both non-cognitive and cognitive in nature. However, no distinction is made between aspects that are social or behavioral in nature; rather, they are closely tied to each other and are therefore, throughout the thesis, referred to as 'social-behavioral' aspects.

It is important to note that there is no clear-cut boundary between such social-behavioral aspects and cognitive aspects since they could be closely tied to each other (Levin, 2011). Rather, it is recognized that being motivated for learning, being interested in schoolwork and taking responsibility would also facilitate learning. Nonetheless, the distinction is made between what is tested in different forms of achievement tests and the knowledge and skills stated in the curriculum, and social and behavioral aspects such as taking responsibility, which are not to form a part of grades or achievement tests. However, such social-behavioral aspects are not directly investigated in the present thesis. Rather, they are believed to be important aspects of schooling and reflected in grades both directly and indirectly.

Grading systems

Grades are summative assessments of the knowledge and skills stipulated in the curriculum that students have acquired. Assessments of knowledge always have to be done with reference to something; there is no absolute scale of knowledge, and grading systems differ with respect to the point of reference used. Two different grading systems can primarily be identified in the Swedish context; the norm-referenced and criterion-referenced systems. Norm-referenced grades have the distribution of grades in the norm-group as their point of reference, the primary function being to rank students for selection. Criterion-referenced grades have goals and criteria in the curriculum as the point of reference, the primary purpose being to provide information about the knowledge and skills acquired in relation to these goals and criteria. This implies that the two systems differ with respect to the interpretation of the

results of the different modes of assessment. It is important to note however that the tests or different types of classroom assessments underlying the grades can be identical, even though different functions of the grading systems have been used at different times.

Norm-Referenced grades

Norm-referenced measuring implies that a score of an individual is meaningful only in comparison with other individuals' scores. It is the norm-group which constitutes the basis for comparison, hence the term 'norm-referenced' (Popham & Husek, 1969). The primary purpose of norm-referenced measurement is to make relative comparisons among individuals for selection purposes. The assumptions underlying this type of measurement are that individuals differ from each other on different characteristics and "that a measure obtained on any physiological or psychological variable for one individual can be reported relative to a distribution of measures of that variable for other people" (Taylor, 1994, p. 237).

The decision to implement a norm-referenced grading system was taken in 1949, the primary purpose being to rank students for selection (SOU 1942:11). A scale with seven letter grades was used, which should follow the normal distribution. However, from 1962 onwards students were to be graded on a scale from 1-5 in compulsory school (Lgr62). The distribution of the grades was based on the assumption that performance and abilities of students follow the normal distribution curve, and the different grades were to be given to a certain percentage of the students, whereas the grade 3 was to be given to the majority (Lgr62). The fixed percentages were with the implementation of a new curriculum in 1980 (Lgr80) removed and replaced by a recommendation of 3 as the average grade in compulsory school.

Frits Wigforss (SOU 1942:11) was the instigator of the reform of the grading system, the primary purpose being to make grades comparable and fair when used for selection from elementary school (folkskolan) to secondary school (realskola). A background for this was that the tests which were used for selection were heavily criticized and were to be replaced by grades as selection instrument. However, the grades in elementary school were neither reliable nor comparable since norms for grading were not used consistently in different schools and classrooms (SOU 1942:11). If grades were to be used as selection instruments they needed to be based on a reliable and comparable

grading standard. Thus, the commission proposed a norm-referenced grading system where standardized tests were to be used as a tool for achieving comparable grades. The aim of these tests was to achieve fair and comparable grades through providing information to the teacher about the level of performance of the class.

Norm-referenced grades in upper secondary school were based on the same principles as in compulsory school. In the 1960s commission (SOU 1963:42) it was suggested that the absolute grading system in upper secondary school should be replaced by a norm-referenced grading system with 5 steps, identical to that in compulsory school. Central tests, similar to the standardized tests, were proposed to function as an aid for grading. The new rules for grading were implemented in the curriculum for upper secondary school in Lgy 70.

The implementation of norm-referenced grades clearly had a societal perspective and there was a need for making sure that selection was fair and accurate, which also was emphasized by the 1940 commission. The implementation of norm-referenced grades has, in at later times, been characterized as a process with democratic aims, implying that having comparable grades and a fair selection based on scientific grounds, would allow for a more democratic way of selecting students to subsequent educational levels (Husén, 1986). There was a clear aim to have selection based on students' abilities, rather than on economic and social predispositions (SOU 1945:45).

The purpose of selection was emphasized by Wigforss and the Grade Commission, even if the motivational and informational purposes were also important (Andersson, 1991). However, the somewhat competitive element built into the system was not seen as detrimental to motivation; rather the opposite was the case and a competitive element was encouraged at a class level (Andersson, 1991). Regarding the informational structure, it was emphasized that grades should measure both knowledge and skills and function as guidance for the school and for parents.

Even though the norm-referenced grading system was in use in Sweden for a long period of time it was heavily criticized on several grounds, not least that it failed to fulfil motivational and informational functions. Grades were criticized for not giving information about the level of a student's performances (they simply showed how one student performed in relation to others (Gustafsson, 2006)) and for encouraging competition among students

rather than cooperation. In particular the norm-referenced grades were criticized because of the fact that teachers seemed to misunderstand the theory of normal distribution and implemented it within the class rather than on the population studying the subject (SOU 1977:9). However, although this was most certainly different among different teachers, it was often voiced as an argument against norm-referenced grades. Further, within the norm-referenced grading system, it was presupposed that all students in the country studied the subject in question, since only then was there a basis for assuming a normal distribution, an assumption clearly problematic for certain subjects and certain courses (Richardson, 2004; SOU 1977:9). In particular this became problematic in upper secondary school due to selected samples studying different tracks and subjects.

Standardized tests

In order to achieve comparability among grades, standardized achievement tests (from here onwards: standardized tests), were implemented to adjust the grading. The standardized tests were to regulate the grading on a class level rather than on an individual level. Thus, the teacher had great freedom in ranking individuals within the class, even though it might be argued that, in some cases, the tests had a strong controlling function. With the implementation of Lgr80, the purpose of the tests was expanded to include the diagnosis of knowledge and skills, a concretization of the curriculum, and a base for research. The tests were to correspond with the grading scale, but, as mentioned above, the fixed percentages were removed with the implementation of Lgr80, and grade directions were only provided for the grade 3, 'higher than 3' and 'lower than 3' (Ljung, 2000).

Standardized tests were provided in Swedish, English and mathematics. However, since students were to choose between general and advanced courses in the subjects mathematics and English, the standardized tests were provided in two versions for the different courses. The course the students took defined the norm-group that an individual student's achievement was to be compared with, indicating that the normal distribution assumption was problematic even in compulsory school. Although the scoring of the tests was carried out by the teacher, they were standardized to a high degree, thus limiting the influence of subjectivity in the grading process. In 1963 standardized tests (called 'central tests') were implemented in upper secondary

school, having the same purposes as the tests in compulsory school (Ljung, 2000).

Criterion-Referenced grades

Glaser (1963, p. 519) asserts that criterion-referenced measurements "depend upon an absolute standard of quality", implying that the individual's performance is compared to a standard or a criterion. The assumption underlying criterion-referenced measurement is that there is a continuum of knowledge where each level of knowledge can be identified and used to describe the specific tasks a student should perform in order to reach that level. Criterion-referenced measurement is primarily designed to give information on the degree of competence/knowledge an individual has attained in comparison to some sort of criterion, not the degree of knowledge of other individuals (Glaser, 1963). While criterion-referenced measurement is mainly used for decision-making regarding whether a student has mastered a certain skill, or for evaluating educational programs (Glaser 1963; Popham & Husek, 1969), norm-referenced measurements are primarily used for rank-ordering. However, even though criterion-referenced measurement does not have a specific competitive selection purpose, such measures are indeed often used to rank individuals, as in the case of the criterion-referenced grades in the Swedish school system.

A criterion-referenced grading system was implemented in compulsory and upper secondary school in 1994 as part of the new curricula for compulsory school (Lpo94) and upper secondary school (Lpf94) (Ds 1990:60). Students were to be graded in Grade 8 and 9 in compulsory school and in upper secondary school after each completed course, on a scale with four different grades; not pass (IG), pass (G), pass with distinction (VG) and pass with special distinction (MVG). The not passing grade was not given in compulsory school, instead the classification of "not yet reached the goals" (EUM) was used.

The informational function was an important aspect in the transition to criterion-referenced grades, which should be better equipped than norm-referenced grades to give information about the students' development of knowledge. Another important purpose was that they should be able to be used as an evaluation of the school. The selection function was not considered in the construction and the implementation of the grading system,

but is indeed an important function of criterion-referenced grades. This is particularly, since currently competition for upper secondary places is on the increase due both to the introduction of independent schools, and the free school choice. This implies that students can choose whichever school they find suitable, leading to increased competition for study places at high-status schools and high-status study tracks (Tholin, 2006). There is also a need to rank students for selection to higher education.

Within the criterion-referenced grading system, the students' knowledge and skills are measured in relation to pre-specified criteria in the syllabus and curriculum (Lpo94 and Lpf94). The goals describe which abilities and which knowledge within every subject the student should develop and the grade criteria describe the levels that need to be achieved for each grade level. (National Agency for Education, 2009a). However, the curricula for compulsory school and upper secondary school are not particularly explicit in describing the content and subject matter for each subject, implying that only goals and criteria or performance standards were given for each subject, and not content standards. Rather, a broad degree of freedom is given to schools in choosing content, methods and material to achieve the goals. The teachers are to interpret the goals and criteria in the syllabus and curriculum which, to a high degree, presupposes that teachers are proficient in practices of assessment and grading.

The criterion-referenced grading system has also been afflicted by problems and has been criticized, for example, for not providing comparable grades and for suffering from grade inflation (Cliffordson, 2004a; Wikström, 2005). It is also criticized for being unclear and difficult to implement in the grading process (Kroksmark, 2002). Consequently, it has at the present time been extended with a somewhat more differentiated scale and a new curriculum which was implemented in 2011 (Lgr11). The system is based on the same principles as its predecessor, and its selective function and predictive validity are still important aspects. The scale for grading is A-F, where A is the highest grade and F is a non-passing grade. However, the main difference between the two curricula is that with Lgr11 goals, performance standards and content standards are given. Moreover, the intention of the curriculum is that the goals, criteria and content should be clearer (prop. 2008/09:87).

National tests

The national-tests have, in accordance with the standardized tests, the function of supporting equity in assessment and grading. An additional purpose is that they should be a base for the evaluation of whether educational goals have been reached at school and national levels. They also serve as a way of explicating and concretizing the goals and grade criteria for every subject, and for assessing the student's level of achievement. Their primary purpose is not to rank students, but rather to assess whether the student has reached the goals in the curriculum.

National tests were, up until 2010, provided in Swedish, English and mathematics in Grade 5 and 9. National tests are also provided in upper secondary school. As in compulsory school the national tests function as a support for grading and form a basis for the analysis of how the goals in the curriculum are achieved on school- and national levels. The functions of national tests are heavily relied on in the Swedish school system, and with the implementation of Lgr11 the national tests have been expanded to encompass earlier school grades and more subjects. However, even though the national tests form an important base for grading, they should not be the sole basis for the final course grade.

The national tests cover different abilities in each subject and the abilities tested correspond to the goals for each respective subject. However, not all subject goals are tested in the national tests. Something that is unique to the Swedish school-system is the fact that teachers carry out all the scoring of the national tests and, in accordance with the assignment of grades, the scoring relies heavily on the teachers' professionalism and unique competence (National Agency for Education, 2012a). It could be argued that due to the reliance on teacher scoring there are threats to the objectivity of tests and aspects which are irrelevant could influence the scoring. However, in order to achieve comparability, rigorous grading criteria and plentiful student examples are provided along with the tests. Teachers are also strongly recommended to cooperate with colleagues in the grading process (National Agency for Education, 2004a).

Research on the National Tests

In evaluations of the effects of the national tests it was found that they function well in supporting the teachers in their grading (National Agency for

Education, 2012a). However, in contrast to these findings, Gustafsson, Erickson and Cliffordson (2014) identify a number of issues indicating that in some subjects the national tests offer limited support for teachers' grading. For example, the number of students getting a non-passing grade on the national tests in mathematics is significantly higher than the number of students getting a non-passing subject grade. In most subjects there is also a high variability in test-grades from one year to another.

Teachers' assessments of the national tests have also been heavily criticized by The Swedish Schools Inspectorate (2012). The results show that there are differences between the teachers who assessed the tests and the external assessors who were re-assessing the tests. These differences were mainly found in the essay part of the English and Swedish tests and, for the most part, were negative, implying that the re-assessors awarded a lower grade than the student's own teacher. These discrepancies could indicate problems with the validity and reliability of the national tests in that the results of the tests depend on who the teacher is (Schools Inspectorate, 2012).

However, the re-assessment of the national tests has been criticized by for example the National Agency for Education (2012b). First, they point out that the agreement between the original assessments and the School Inspectorate's assessments is in general quite high. The discrepancies found mainly concerned the essay-tests in Swedish and English where, in compulsory school, the agreement is 56 per cent for the Swedish test, and 62 per cent for the English test. The National Agency for Education argues that the School Inspection has not taken into account research on inter-rater reliability, where agreement between 40 and 70 per cent on essay-type tests is considered high (e.g. Brennan, 2006). In the re-assessments there are more often negative discrepancies than positive. However, stricter judgments do not imply that the judgments are more correct. Moreover, in a third re-assessment of the tests with the largest discrepancies, an equal number supported the original assessment. It was also shown that a different scale was used in the re-assessments, which could affect results in borderline cases. Furthermore, there is no justification for why the re-assessors should be better in assessing the tests than the original teachers (National Agency for Education, 2012b).

In a study by the National Agency for Education (2009b) the agreement between raters for the Swedish writing test was 54 percent, but the result varied depending on how much training the raters had. The agreement for the English writing test ranged from between 86-93 percent. Still, there are

differences in the ratings between teachers and the re-assessors and there are indications that teachers might be leaner in their assessment. However, many of the re-assessors from the Schools Inspectorate also argue that they follow the assessment-directions more strictly than when they are assessing their own students (National Agency for Education, 2012b).

The re-assessments have also been criticized by Gustafsson and Erickson (2013) concerning the degree of representativeness of the sample of teachers chosen to reassess. The sampling design of schools was also criticized, given that each school only was represented by one subject leading to school differences also being confounded with teacher differences. Gustafsson and Erickson (2013) also point out that there are indications that the re-assessing teachers might be negatively biased in their assessments because they have interpreted their assignment in such way that they become harsher in their assessments.

National tests have also been shown to measure different dimensions of abilities. Åberg-Bengtsson and Eriksson (2006) identified different dimensions in the national tests. Using two-level structural equation modelling they identified on the within level a broad structural factor which was related to the mathematics test measuring basic skills, the English test measuring receptive skills, and to the Swedish reading comprehension test. The structural factor was distinctly separate from the listening/creative factor, indicating that being able to argue using verbal language is distinctly different from handling structural information as represented by, for example, mathematical symbols and linguistic features. Moreover, a factor representing communicative mathematic skills was identified, related to tests measuring oral communication and problem-solving. The study indicates that national tests measure different dimensions of abilities, which are related or not related to each other in different manners (Åberg-Bengtsson & Eriksson, 2006). Eklöf and Nyros (2013) also show that social-behavioral aspects, such as perceived importance and invested effort and motivation, had a positive relation with test results. These results indicate that national tests may reflect both cognitive and social-behavioral dimensions.

The studies cited above indicate that the national tests could suffer from reliability deficiencies, which in turn may lead to results not being valid. Inter-rater reliability is certainly an important aspect to consider when considering test reliability. The higher the inter-rater reliability, the better the test, all other things being equal (Stemler, 2004). However, only focusing on reliability

aspects in a test would lead to test tasks which are easy to assess. It would result in an underrepresentation of the more qualified knowledge and skills stipulated in the curriculum and, ultimately, a poorer operationalization of the goals and criteria in the curriculum (National Agency for Education, 2009b). A strength of the tests is that they are constructed from the goals and criteria in the curriculum (also measuring the more complex goals stated there) and contribute to the concretization of these goals (SOU 2007:28). This is important since the tests are a part of the governmental steering of the school (National Agency for Education, 2012b). Still, the goals and grading criteria in the curriculum, and also the grading criteria for the national tests, are open for interpretation which can result in differences in how, for example, the teaching is conducted and also how assessment of both the national tests and of other assignments are executed by teachers. However, nationally representative studies on the practical and pedagogical function of the national tests and how they are valued by teachers and students indicate that they have a strong legitimacy (National Agency of Education, 2004b).

It is important to note that even though the tests in compulsory school have similar purposes and function as a support for grading, they also differ in their construction. While the tests used within the norm-referenced grading system aimed to create standards in order to categorize students according to a set of grades, the national-tests used within the criterion-referenced grading systems aimed at testing the extent to which students fulfil pre-formulated standards. Throughout the thesis when referring to both the standardized tests and the national tests the term 'achievement tests in compulsory school' will be used.

3 Dimensionality of grades

Grades are summative assessments of the knowledge and skills that a student has acquired. Even though other goals are stated in the curriculum as well, a grade should be an assessment of the students' actual competence in a certain subject, implying that grades do only measure a part of the goals set up in school (Gustafsson, 2006). There are also explicit functions of grades which grading systems have to be evaluated against. These explicit functions are: a) the informational function – i.e. to give information about the level of knowledge and skills a student has acquired to the student, to parents and to the school, b) the motivational function – i.e. to function as a motivation and an incentive for students' future studies, c) the selective function – i.e. they should be able to rank-order students for selection to subsequent levels in the educational system and, to some degree, predict how a student will be able to perform in the future.

However, in previous evaluations of grades there have been indications that there are implicit functions of grades which are not stated in the curriculum or other steering documents. These implicit functions exist, for example, in unspoken rules and expectations about how students should behave and in implied rules concerning the attitudes towards knowledge one should have in order to receive a certain grade. These implicit rules and expectations stand in sharp contrast to the construct that grades are supposed to measure, i.e. knowledge, skills and abilities as they come to be portrayed in the curriculum (Klapp Lekholm, 2009), as well as being in sharp contrast to the explicit functions.

A great amount of research also shows that grades do not only measure knowledge and skills, but that they are multidimensional measures. Grades seem to encapsulate not only knowledge and skills, but also a plethora of other aspects (e.g. Alexander, 1935; Andersson, 1998; Bowers, 2011; Gustafsson & Balke, 2003; Klapp Lekholm & Cliffordson, 2008, 2009). In a seminal study by Alexander (1935) several factors in school grades were identified, of which one was named X. The X-factor was viewed as exerting an important influence for success in all school subjects, and interpreted as representing persistence and determination. Research has, in accordance with

the early results of Alexander (1935) shown that both norm-referenced and criterion-referenced grades have a multidimensional structure (e.g. Gustafsson & Balke, 1993; Klapp Lekholm & Cliffordson, 2008). Gustafsson and Balke (1993) investigated the dimensional structure of norm-referenced grades. They identified latent factors reflecting different subject areas but also identified a general school achievement factor. This general factor influenced all grades and was related to students' cognitive ability. However a part of the variance in the general school achievement factor could not be explained. It was hypothesized that this unexplained variance was related to behavioral and motivational aspects, but also quite substantially to verbal competencies (Gustafsson & Balke, 1993). Andersson (1998) further investigated the dimensional structure of norm-referenced grades. A general school achievement factor was identified which was related to all grades. The interpretation of the school achievement factor was however not entirely clear and it was hypothesized that it reflected motivation and adjustment to the school system. Moreover, neatness and manners were believed to influence the general factor.

The dimensional structure of criterion-referenced grades has been investigated by Klapp Lekholm and Cliffordson (2008, 2009). Grades were shown to be multidimensional, reflecting subject-specific dimensions measured by tests and grades, and a common-grade dimension which cuts across grades and teachers (Klapp Lekholm & Cliffordson, 2008). The common-grade dimension could primarily be related to students' interest and the parents' commitment to their children's' studies (Klapp Lekholm & Cliffordson, 2009). It was also discussed that these discrepancies between tests and grades could reflect the complexity and the dilemmas that teachers face in managing the classroom and motivating students – i.e. as a moral dilemma where there is a pressure of being objective and, at the same time, having a subjective relation with the student (Klapp Lekholm & Cliffordson, 2008).

Similar results concerning dimensionality have been identified by Bowers (2008, 2011). Grades were found to consist of both cognitive and social-behavioral dimensions. Drawing on standardized achievement tests and teacher-assigned grades, Bowers (2011) identified one dimension related to knowledge and skills as measured by standardized achievement tests, and one dimension which was different from the knowledge dimension, and which was argued to be related to social-behavioral aspects. It was hypothesized that this dimension could be related to aspects such as effort and engagement and

"thus ultimately their success at being schooled" (Bowers, 2011, 153). Bowers (2008) also suggests that grades should be reconsidered as valuable measures of both knowledge and skills, and of behavioral and social aspects important for school. It is suggested that grades measure both knowledge and skills and the "student's ability to perform well at the process of schooling" (p. 623).

Accuracy of teachers' assessments

The research on the dimensional structure of grades indicates that grades are multidimensional measures. This dimensionality has partly been related to teachers' grading practices. Similarly, the quality in teachers' grading has typically been measured by the discrepancies between different forms of standardized tests on students' achievement and grades. While grades traditionally have been regarded as subjective measures standardized achievement tests get to constitute the criterion measure which have been regarded objective. Research on the quality of teacher assessments has traditionally shown that teachers' assessments tend to correlate moderately to highly with tests of students' actual achievement. Even though the correspondence tends to be fairly high, there is unexplained variance which cannot be explained by student achievement (Südkamp, Kaiser & Möller, 2012).

The National Agency for Education (2007) has investigated the correspondence between grades and national tests in 9th grade. The results showed that for most students the grade on the national test and the final grade were identical, the correlation being about .80. However the results also showed that there were substantial differences between teachers. It was significantly more common that the students got a higher grade compared to the grade on the national test in mathematics and Swedish. In English there were as many who, in comparison to the test result, got higher grades as who got lower grades. It was primarily students who did not reach the goals and got a non-passing grade on the national test that received a higher subject grade (a passing grade).

Moreover, differences between teachers were found in an early review on the accuracy of teacher judgments by Hoge and Coladarci (1989). Typically they found that the teacher judgments were accurate and had a median correlation with standardized achievement tests of .66. But as in the case of the study by the National Agency for Education, there was great variability among teachers. Still, since teacher judgments encompassed a wider array of

outcomes they concluded that the teacher assessments had a high validity. In a later review by Harlen (2005) similar results were found, implying that teacher assessments in general are consistent with test results, but that there was great variability between teachers.

Südkamp, Kaiser and Möller (2012) have in a meta-analysis investigated the correspondence between teachers' judgment of academic achievement and actual student achievement as measured by tests. They found that teacher accuracy as defined by the correlation between teachers' judgment and students' test performance, was fairly high, with a correlation of .63. However, they point out that the correlation is far from perfect and there still remains unexplained variance. In a Swedish context Johansson, Myrberg and Rosén (2012) investigated the correspondence between teacher judgments of students' reading skills and the achievement on the Progress in International Reading Literacy Study (PIRLS). They found, much in line with the research reported above, that there was a fairly high correspondence between test results and teacher judgments within classes, but not between teachers.

Teachers' grading practices

The accuracy of teachers' assessment is important to investigate since the assignment of grades relies heavily on the expertise of the teacher. The teacher has to interpret goals and criteria in the curriculum and syllabus for each course and each subject, and it is assumed that the teacher can make fair and reliable assessments. The Swedish Education Act (2010) states that there shall be educational equality between schools irrespective of type and where in the country the education is provided. This implies that all students, regardless of whom they are, should be provided with adequate and appropriate schooling. This also implies that aspects which are irrelevant to the construct of grades – such as social-behavioral aspects, gender or socio-economic background – should not influence teachers' assessments. However, research on the dimensionality of grades and on the accuracy on teacher judgments indicate that there might be threats to both the reliability and the validity of grades.

Gender

There is research which indicates that there is an influence of irrelevant aspects to grades. Research has identified that, in recent decades, girls have received successively higher grades than boys. This advantage for girls is

evident on all achievement levels and in all groups defined by social or cultural background (Wernersson, 2010). Girls also seem to gain higher grades in almost all subjects (Willingham & Cole, 1997).

Svensson (1971) showed that girls get higher grades than boys relative to their achievement on standardized achievement tests. Traditionally these discrepancies have been interpreted in terms of girls' higher grades being unjustified in relation to their achievements on tests (e.g. Emanuelsson & Fishbein, 1986). However, in an analysis of changes in cognitive test results between 1961 and 2005, Svensson (2008) found that while both groups have positive developments, girls have higher means on both the verbal and the spatial test. In light of these results Wernersson (2010) argues that the girls' grades are instead now more in line with their results on cognitive tests. However, when investigating both the original assessments and the re-assessments of the national tests made by the School Inspectorate, Hinnerich and Vlachos (2013) found that the teachers' test-scores were lower for boys than for girls in the original assessments compared to the re-assessments.

The National Agency for Education (2006) found that some of the differences in grades can be explained by the fact that girls put more time and engagement into their school work, which might be connected to beliefs about male and female identity. However, Nycander (2006) argues that girls' higher grades in relation to national tests can partially be explained by girls being compensated for their supposedly subordinate position in the classroom, and that they are rewarded for their inclination to be helpful.

There is certainly no single factor which can explain gender differences in grades, and theories attempting to explain these differences take different directions. Furthermore, Klapp Lekholm and Cliffordson (2009) have shown that the advantage for girls in the common-grade dimension is mediated by interest. Such aspects could of course affect grades indirectly by resulting in increased learning, but also directly by being aspects which are recognized as important in the classroom, and thus for grades.

Educational background

There also seem to be differences in grades which are due to students' socio-economical background (SES). Coleman et al. (1966) found that social background was one of the most important influences on achievement. Even

though many years have passed since Coleman's study, students' SES still seems to be important.

The concept of socio-economical background is often a composite of cultural, economic, social and educational capital. However, Yang (2003) found that the different dimensions of SES influenced achievement differently, and that the SES concept is multidimensional. When investigating the different aspects of the SES concept separately, economic capital did not have a positive impact on achievement. Rather, the parents' education was the most important aspect for school achievement. The importance of the educational environment in the home is also emphasized by Bloom (1976). Similarly, in studies like PISA, the number of books at home stands as a proxy for SES, where a large number of books in the home indicates a family environment that promotes education and academic effort (Woessmann & Peterson, 2007).

Students with higher home educational backgrounds tend to be high achievers in school (Hattie, 2009; Woessmann, 2004) and to choose more theoretically-oriented tracks (Reuterberg & Svensson, 2000, Svensson, 2001). In addition, there are indications that the impact of parents' education for students' results is constant (e.g. Gustafsson & Yang Hansen, 2009; Böhlmark & Holmlund, 2011). However, the National Agency for Education (2012c) detected a slight increase in the importance of parents' education when using a finer-graded scale.

The influence of educational background on grades and tests could be a result of the fact that these students have more support from the home environment with the school work, which increases learning and leads to higher achievement in school. It could also be argued that students from homes with a higher educational background are more aware of codes and conduct in school (Björnsson, 2005) which could thus influence grades, even if these aspects should not be of importance in the grading process. In addition, the school has a compensatory role, implying that it should compensate for the influence of social background.

Research on the dimensionality of grades reveals a complex pattern of the influence of educational background on the different dimensions of grades (Klapp Lekholm & Cliffordson, 2008, 2009). Apart from the differences due to educational background on grades, there are additional differences in the common-grade dimension to the advantage of students with lower educational backgrounds. These differences were interpreted as a compensatory grading practice employed by teachers, implying that students

are compensated for their disadvantaged situation by teachers considering social-behavioral aspects to a greater extent in their grading. It was argued by Klapp Lekholm and Cliffordson (2008, 2009) that considering such social-behavioral aspects in the grading process might be an attempt to motivate students, meaning that teachers are concerned with the consequences of grades. Partial evidence of a compensatory grading mechanism is also shown by the fact that students with a lower educational home background tend to be overrepresented in the group of students who gain a passing grade despite the fact that they do not reach all the goals for a passing grade (National Agency for Education, 2012c).

In contrast to the findings of a compensatory mechanism, Hinnerich and Vlachos (2013) found in their investigation of the different assessments of the national tests that the teachers' test-scores were lower for students with lower home educational backgrounds in the original assessments. These results rather suggest that students who tend to have relatively weak study results are disadvantaged by teachers in their assessments. However, the effects were quite small and concern test-scores not grades, and are in conflict with what has been found elsewhere (e.g. Klapp Lekholm & Cliffordson, 2008).

Social-behavioral aspects

Results have also shown that teachers tend to take social-behavioral aspects into account when assigning grades (e.g. Brookhart, 2013), implying that teachers award grades both on the basis of knowledge and skills and on a multitude of other aspects (Brookhart, 1991; Cizek et al., 1995; Cross & Frary, 1999; McMillan, 2003; Randall & Englehard, 2010). Brookhart (1991) refers to this as a "hodgepodge grade of attitude, effort and achievement" (p. 63). Studies on teachers' grading practices have found that, when assigning grades, teachers tend to consider both achievement and aspects which are not related to achievement, such as social-behavioral aspects, (Cross and Frary 1999; McMillan 2001; McMillan, Myran, and Workman 2002; Randall and Englehard 2009a, 2009b, 2010).

In Brookhart's (1993) analyses of teachers' grading practices Messick's (1989) validity theory was used. It was found that teachers tend to argue that taking account of effort and other social-behavioral aspects in their assignment of grades was important in relation to treating their students fairly and since they are concerned with the consequences of grades. Thus, what is

recommended in the grading standards and teachers' concerns about social consequences come into conflict. Additionally, it was found that this type of grading practice was motivated by its importance for the development of students' self-esteem. Further, it was argued that the teacher functions both as a judge and an advocate for the student and that they have to make compromises between these two roles, which results in the inclusion of irrelevant aspects.

Annerstedt and Larsson (2010) investigated grading in physical education in Sweden. They found that the grading could be questioned with respect to equity and fairness and that the grading was somewhat arbitrary. For example, the teachers in physical education considered behavior as an important aspect when assigning grades, and teachers also seemed to have internalized the criteria of a certain grade. However, these internalized aspects were not based on knowledge and skills.

Moreover, students' individual development plans have been shown to be influenced by understandings of the "perfect student". An individual development plan is a document which is meant to function as a support for students' learning and it should consist of two parts; one which is an assessment of knowledge and skills, and another which consists of aims which the student should work towards. As shown by Mårell-Olsson (2012), there is much focus on attitudes and behavior in the development of an individual development plan. This indicates that students are expected to conform to the implicit expectations from school, and in the individual development plan that knowledge and skills are connected to the behavior in school (Mårell-Olsson, 2012).

Researchers, especially within the psychological domain, also indicate that halo effects or halo errors are a part of subjective assessments (Dennis, 2007). Indeed grading can in some respects be classified as a subjective assessment of students' achievements. Halo-error is classified as a cognitive bias in the assessment of performance (Dennis, 2007). According to the results and definition by Nisbett and Wilson (1977) the halo effect implies "that global evaluations of a person can induce altered evaluations of the person's attributes, even when there is sufficient information to allow for independent assessments of them" (p. 250). This implies that if we like a person the assessment of other attributes which are not known will be affected by this overall impression. The contrary scenario is also true (Nisbett & Wilson, 1977). Investigations of halo effects in educational settings are rare. However,

Dennis (2007) investigated the halo effects on grades awarded by two raters on undergraduate projects. There were halo effects on the grades awarded and the effects did not differ with respect to how regular the contact with the students was. This implies that halo effects could be present in teachers' grading causing influence of irrelevant aspects in teacher-assigned grades.

The decision-making concerning grades relies primarily on classroom assessment. In classrooms, many factors are in play that affect such assessments. McMillan's (2003) theory on classroom assessment takes account of the myriad of these factors and how they can affect teachers' decision-making. They include the teacher's own beliefs, expectations and values, as well as external factors. Hence, teachers are faced with a complex situation where, on the one hand, they are both have to balance classroom reality in motivating students, their own values and attitudes towards teaching and learning, and, on the other, carrying out objective and fair assessments (Klapp Lekholm, 2010). Brookhart (1993) indicates that this complex classroom reality might be reflected in teachers' grade assignment, which could imply that grades reflect construct-irrelevant variance (Messick, 1989) which, in a long-term perspective, would have consequences on several levels both for individuals and institutions.

4 Cognitive abilities

Spearman's two-factor theory of intelligence can be regarded as the starting point for intelligence theories based on factor analytical approaches. His two-factor theory could account for the correlations that were found between different measures of cognitive performance (Shneider & McGrew, 2012). Spearman found that both measures of achievement and different measures of reasoning and sensory discrimination were positively correlated. From this he concluded that the correlations between these measures could partly be determined by a general factor of intelligence, *g*. Spearman's findings were challenged by Thurstone who rejected the notion of a general factor and instead suggested that there were several distinct abilities and that the correlations between these were not indicative of a general factor.

However, to this date, there is ample evidence that a hierarchical structure of intelligence is able to reflect the complexity of human intelligence. The hierarchical model of intelligence as defined by Horn and Cattell (1966) identifies factors of differing generality at different levels. A large set of narrow factors are at the lowest level of the hierarchy, and a small set of broad factors at the second level. On the second level one can identify the broad factors *Gf*, a general de-contextualized ability to reason and solve problems in novel situations; and *Gc*, which represents knowledge and skills acquired through education and experience in several domains (Cattell, 1963; Horn & Cattell, 1966). Cattell did not define a third-order general factor of intelligence and, instead, Spearman's *g* was broken down into *Gf* and *Gc*. However, in an extensive empirical study by Carroll (1993), an extension of the Horn-Cattell *Gf-Gc* model was defined with a third order general intelligence factor at the apex of the model. A representation of both Cattell-Horn's and Carroll's models integrated is referred to as the Cattell-Horn-Carroll (CHC) model.

Schneider and McGrew (2012) describe the CHC model as a taxonomy of cognitive abilities, intended to theoretically describe how people differ in their cognitive abilities. The CHC model is empirically supported and is regarded as the most comprehensive model of the structure of human intelligence (McGrew, 2009). However, there are still controversies concerning the definition of a general factor. While for example Carroll (1993) argues that *g* is the cause of the positive manifold, i.e. the positive correlations between tests of

mental ability, Horn and Blankson (2012) argue that almost all tests of cognitive ability correlate positively and, if such a general ability exists, it will be weak, hard to detect and hard to interpret. However, a solution to this controversy has been identified by Gustafsson (1984) who argues that the *g* factor, as it is perceived by Spearman, is much in accordance with the *Gf* factor as it is perceived by Horn and Cattell (1966). In several empirical studies it has been found that *g* and *Gf* can be defined as one and the same factor since unity between these has been identified (e.g. Gustafsson, 1984, 1988; Valentin Kvist & Gustafsson, 2008).

Investment theory

While the structural part of the human intelligence has gained substantial attention, Cattell's (1971, 1987) developmental Investment theory has not received similar recognition. Investment theory concerns the development of intellectual abilities and states that there is a causal relationship between *Gf* and *Gc* where *Gf* is considered as a general ability influencing the development of knowledge and skills. While there are a number of studies testing the hypothesized relations in Investment theory, the patterns of results are somewhat unclear. However, the perfect relationship between *Gf* and *g* identified by Gustafsson (1984) and Valentin Kvist and Gustafsson (2008) could give indirect support for the Investment theory:

> This perfect relation between *Gf* and *g* may, according to the Investment theory, be interpreted to be a consequence of the ubiquitous involvement of *Gf* in acquisition of knowledge which implies that *Gf* will be a source of variance in each and every task which requires learning of new knowledge and skills. These results therefore provide indirect support for the Investment theory" (p. 9).

Cattell's (1971, 1987) Investment theory can be considered as an extension of the structural *Gf-Gc* theory or the developmental part of the *Gf-Gc* theory which describes the causal relationship between *Gf* and *Gc*. Stated more mundanely, the theory was intended to provide an answer as to why some people know more than other people, positing that people with high level of *Gf* have fewer limitations in acquiring new knowledge than people with lower levels of *Gf* (Schneider & McGrew, 2012). This implies that individual differences in *Gc* depend on the level of *Gf* and, due to the involvement of *Gf*

in all knowledge acquisition, *Gf* will be reflected in all tasks reflecting knowledge.

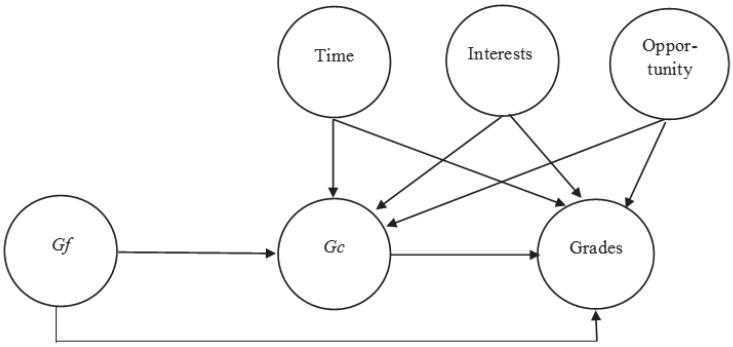

Figure 1. Adaptation of Cattell's Investment model

Gf is formulated to be a general, relation-perceiving ability connected with the maturation of the brain which is associated with genetic factors. This general ability can be applied to any sensory, motor or memory area and determines the rate of learning different tasks. The influence from *Gf* is particularly strong in learning of areas which demand understanding of complex relations, such as reading, arithmetic and reasoning. Hence, learning of such complex areas will depend on the level of *Gf*. Knowledge and skills are then developed through practice and experience. These acquired abilities are labeled *Gc*, which are influenced by the general ability *Gf*. As is evident in Figure 1, Cattell (1987) also recognized that much learning comes about due to many non-ability factors. These non-ability factors are involved in the investment people out into learning, meaning that people invest, for example, time, effort and interests into the acquired abilities *Gc*. The acquired abilities also depend on the opportunities to learn, meaning that without the opportunity to learn there would be no return on *Gf*.

This implies that all these factors are invested into learning, and that knowledge and skills are developed through practice and experience. The acquired abilities – i.e. *Gc* – are influenced by *Gf* which also determines the level of the output of this investment. In turn *Gc* also enables children's learning in school in fields such as reading and writing. These acquired abilities can also be labeled *Gc*. Accordingly, school achievement is influenced by *Gf* and the

previous level of *Gc* as well as the time and interest invested, and the opportunity to learn, both directly and indirectly via *Gc* (Figure 1).

However, the pattern of results in previous studies testing the theory is rather unclear. This might in part be explained by the definition of *Gc*. The difficulties involved in defining *Gc* are recognized by many researchers (e.g. Ackerman, 1996, 2000; Gustafsson & Carlstedt, 2006). While Cattell (1987) defined *Gc* as acculturated learning in a wide range of domains, different researchers infuse different meanings into *Gc*, both as a composite of different learning achievements in different domains and, more narrowly, as verbal comprehension (Kan, Kievit, Dolan & van der Maas, 2011). Postlethwaite (2011), for example, defines measures of *Gc* as "the efficiency with which an individual has learned a variety of material over a long period of time" (p.150.). Research which has emphasized *Gc* as a predictor over *Gf* has also adopted this somewhat broader definition of *Gc* (e.g. Ackerman, 2000; Gustafsson & Carlstedt, 2006; Postlethwaite, 2011). Nevertheless, Kan et. al. (2011) extended the Valentin Kvist and Gustafsson (2008) finding that *Gf* and *g* have a correlation of unity in homogenous samples by showing that the verbal comprehension factor had a relation of unity with *Gc* when the group is homogenous with respect to education. Thus, in culturally and educationally homogenous samples "the interpretations of *Gc* as exposure to education and *Gc* as verbal comprehension intersect" (Kan et al., 2011, p. 297). The homogeneity and broadness of the *Gc* measure seems to play a part in the results when testing Investment theory.

The somewhat inconsistent results in previous investigations of the theory may also in part be a factor of age. While the direct effects of *Gf* on learning exert the largest influence in the early development of a child through the early school years (Cattell, 1987), previous investigations of Investment theory have typically been conducted with adolescent or adult samples (e.g. Ackerman, 2000; Gustafsson & Carlstedt, 2006). This might imply that the developmental period where *Gf* exerts its direct influence has passed and explain why no direct effects of *Gf* would be evident.

Encapsulation theory

Encapsulation Theory, developed by Gustafsson and Carlstedt (2006), can be regarded as an extension of Cattell's (1971, 1987) Investment theory. Encapsulation theory takes its point of departure in results recognizing the di-

minishing importance of *Gf* and the increasing importance of *Gc* in knowledge acquisition. It is argued that these results arise because measures of *Gc* include *Gf* variance, which is predictive of learning and achievement and, in addition, that *Gc* reflects individual differences in knowledge and skills which are of importance for further learning and achievement. According to Encapsulation theory, the individual differences in *Gf* are encapsulated in *Gc*, and all the information from *Gc* is encapsulated in the high school GPA (Grade Point Average). Encapsulation theory has been empirically tested with a sample of 3089 men admitted to a graduate engineering program. In the empirical part, no direct effects of *Gf* on achievement were found (Figure 2). Alternative models were set up but they either did not fit data well, or had only a marginally acceptable fit, while the original hypothesized model had an excellent model fit.

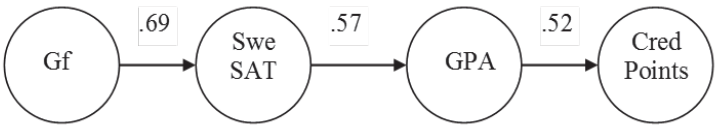

Figure 2. Gustafsson's and Carlstedt's (2006) simplex model.

It was argued that no predictive efficiency is gained by adding a *Gf* measure when measures of *Gc* and prior knowledge are available. This implies that the development of knowledge and skills increasingly relies on *Gc*, and that the influence of *Gf* is primarily indirect and mediated via *Gc* (Figure 2) (Gustafsson & Carlstedt, 2006).

PPIK theory

Another extension of Investment theory is the theory of PPIK (Process, Personality, Interests, and Knowledge) developed by Ackerman (1996) to describe intellectual development in adults. PPIK theory relies heavily on Cattell's (1987) Investment theory. However, instead of drawing the distinction between *Gf* and *Gc*, Ackerman makes a distinction between intelligence-as-process and intelligence-as-knowledge. The intelligence-as-knowledge concept is close to Cattell's (1987) original description of *Gc*, but in line with Gustafsson and Carlstedt's (2006) definition of *Gc*, it is much broader and deeper in scope. It is also emphasized as a predictor. PPIK theory puts more emphasis on domain-specific knowledge and on the importance of personality

and interests in determining the direction of the effort which is invested into the knowledge acquisition and the maintenance of this knowledge acquisition, i.e. of intelligence-as knowledge. The theory has, in a series of studies, been empirically tested (e.g. Ackerman, 2000; Ackerman & Beier, 2006; Beier & Ackerman, 2001; Beier & Ackerman, 2005). These studies confirm the proposal of a broad investment model with *Gc* as the most important factor in the acquisition of knowledge and skills of adults. *Gf*, or intelligence-as-process, is seen as determining the output of the investment. Intelligence-as-knowledge is the result of this investment, but also a result of a series of individual choices. Personality traits determine the direction and intensity of intelligence-as-process and these non-ability factors also drive the maintenance of domain specific knowledge during the adulthood (Ackerman, 2000).

5 Validity

Since grades are used for selection they are high-stakes instruments and can have serious consequences for students. Hence, the validity of grades is an important aspect to consider, and questions concerning the functions of grades can certainly be viewed from a validity perspective. In this area of research Messick's (1989) validity framework has been shown to be particularly useful (e.g. Klapp Lekholm, 2008; Nyström, 2004), because it encompasses both the inferences drawn from assessments, and their consequential aspects. However, Messick's (1989) validity theory does not give particular guidance in practical application in educational settings, which is also beyond the scope of this thesis. It has also been argued that the ever-increasing broadening of validity theories has increased the gap between validity in theory and validity in practice (Wolming & Wikström, 2012).

Moss, Girard and Haniford (2006) argue that the validity theories developed in the educational measurement tradition focus on standardized assessments and thus only gives circumscribed information about student learning and that there is a limited amount of interpretations, decisions and actions which can be supported by the validity evidence provided by such assessments. However, Moss et. al. (2006) also recognize that the resources within the educational measurement perspective can be used when there is a goal to develop a common validity argument which can support the interpretations, decisions and actions which are relevant across individuals and across situations. They also recognize that validity theories within educational measurement are first and foremost developed in order to evaluate the intended interpretations of measurements. In the present thesis the intended interpretations of subject grades are of focal interest and interpretations and decisions which have relevance across individuals, not that is the practical application per se. Issues concerning the intended interpretation of grades, predictive validity and selection are not context-bound; rather they ask for a common validity argument.

Traditional views on validity

Traditionally the validity concept has been divided into different categories of validity: content validity, predictive and concurrent criterion-related validity, and construct validity. Content validity refers to the extent to which the content of a test or a measure covers the content of the domain one measures. Criterion-related validity implies relating the results of an instrument to another external criterion. Criterion-related validity is usually divided into predictive validity and concurrent validity. Predictive validity concerns how well an instrument reflects an external outcome that occurs in the future, for example how well compulsory school grades can predict performance in upper secondary school. Concurrent validity deals with how well an instrument reflects another external measure that occurs simultaneously, or at least in the same time-frame (Messick, 1990).

Early research on different instruments used for selection of individuals to different educational institutions has in the past focused primarily on an instrument's predictive validity. Validity was thus seen as a property of the instrument itself. In Cureton's (1951) seminal text validity was described as indicating "how well the test serves the purpose for which it is used" (p. 621). Accordingly, a test could have many validities depending on the correlation with criterion scores from different areas. Primarily, the view of validity was on relevance, i.e. if the test measures what it was designed to measure. However, during the years after Cureton's (1951) text, validity theories within the educational measurement environment have seen a change in how validity is perceived.

During the second half of the 1900s the conception of validity began to broaden and validation concerned questions regarding the interpretation of an instrument in relation to constructs anchored in psychological theories, rather than questions concerning an instrument's predictive validity. Cronbach and Meehl (1955) introduced the term construct validity, which refers to the empirically based interpretation of an instrument which is also linked to the theoretical framework. They argued that for construct validation it is necessary to employ a theoretical framework suitable for that which is the subject of measurement, but also an empirical framework for how measurement is to take place, as well as specifications for the relations between these networks. Cronbach (1971) also discusses how construct

validity is the overarching form of validity and how all other aspects of validity are subordinate or secondary.

Construct validity

With Cronbach and Meehl's (1955) introduction of construct validity, the conception of validity has changed and broadened. With Messick's (1989) validity chapter the unified validity theory was presented, which gives an even broader view of validity. The unified view on validity implies that it is a scientific inquiry into the meaning of a score. This implies that the validity concept has been expanded to encompass aspects other than merely 'measuring what one intends to measure' and it is no longer a property of a test or an assessment. Rather, validity refers to the interpretations and inferences which are drawn from the results of a test or an assessment and the decisions and actions taken based on these interpretations and inferences.

Hence, validity is not connected to the instrument itself, but to the inferences one makes on the basis of the outcomes. Validity concerns the interpretation of a score and how it is used and which consequences it will have. Messick (1989) also holds that both the empirical evidence and the theoretical perspectives should be consistent with the inferences drawn and the usage of the score. "In essence" he argues, "test validation is empirical evaluation of the meaning and consequences of measurement" (Messick, 1990, p. 2). Hence, the different 'categories' of validity traditionally formulated rather all relate to construct validity since they contribute to the meaning of the score and thus are aspects of construct validity. Further, in the applied use of a test, construct validity needs to be supported by evidence of the relevance of the test to the purpose and the utility of the test in the setting (Messick, 1989).

Messick (1989) also presents a model referring to the different facets of validity. The different facets in the model refer to the evidential basis of test interpretation and test use, as well as the consequential basis of test interpretation and test use. Hence both how a test score, a grade or other type of outcome is interpreted and used, and what evidence and consequences it will lead to, is taken into consideration. It also aims at identifying two major threats to validity; construct underrepresentation, meaning that a measure is too narrow, and construct irrelevant variance, meaning that a measure is too broad, measuring aspects irrelevant to the construct. However, the different

facets are intertwined and both in test interpretation and test use, the evidential and consequential basis needs to be considered.

If we follow Messick's (1989) facets of validity we would first ask if the inferences drawn from a score are valid (from here on grade), which has to do with construct validity; does the grade measure what it was intended to measure? The second part has to do with value implications; how do the interpretations of the grade affect views on knowledge, and what is important to learn? The last two aspects deal with the usage of the grades; can the usage of grades for different purposes can be justified in the relation to the interpretations of the construct of grades, and what social consequences, for example for purposes of selection, attach to the use of grades? Thus, questions that need to be asked are what is the meaning of grades, are they relevant measures in relation to their purposes, are they possible to use in decision- making, and what are the social consequences of using them?.

As stated above there could be several potential threats to the validity of grades, i.e. construct irrelevant variance and construct underrepresentation. If grades reflect such aspects which are not stated in the grading criteria, and if aspects irrelevant to the construct of grades are indeed allowed to influence them, grades would suffer from construct irrelevant variance. Another threat to the validity of grades is construct underrepresentation, which would imply that the grade is too narrow in relation to its construct, implying that greater importance is given to one or a couple of abilities stated in the criterion while others are left out.

If any of these potential threats to validity would be relevant in relation to grades, the inferences drawn could be distorted and the relevance of grades be problematic. It would also lead to social consequences when grades are used for different purposes, such as selection (Allen, 2005). However, Messick (1989) argues that adverse social consequences of using a test or a grade, for example for selection, does not lead per se to an interpretation that the usage is invalid. Only if the adverse social consequences can be connected to sources of invalidity, such as construct irrelevant variance, can one question whether the usage of the test or grade is valid. Consequences of using grades which suffer from construct underrepresentation or construct irrelevant variance would be evident for both individuals, institutions and society.

There are opponents to Messick's (1989) inclusion of social consequences into the validity concept. Some researchers argue that social consequences of using tests or grades are an important issue to consider, but that it should not

be mixed up with the validity concept. However, since grades play such a great part in people's future opportunities, and since they are used for decision-making that impacts on people's life chances, social consequences seem to be an important part of the validity of grades. It is important to note though, that when evaluating the validity of grades, and indeed also the social consequences of using grades, Messick (1989) argues that one has to juxtapose the supposed use of the assessment against other types of assessments supposed to serve the same purpose.

Predictive validity

Messick's (1989) unified validity theory clearly shows that construct validity is the overarching validity, while the other types of evidence are subordinate. Still, as shown by the argument above, these subordinate validity types are also important when considering the validity of a measure. Messick (1989) argues that decisions such as selection are based on the inferences drawn from a specific measurement. Grades and other types of tests used for selection are assumed to predict future school achievement, which implies that predictive validity is indeed an important aspect of validity to consider.

Aptitude and achievement

The kind of solution or type of instrument used for predicting future learning and achievement is a matter of whether aptitude or achievement should be focused on when rank-ordering individuals for selection. This distinction between aptitude and achievement was present already during the development of the first intelligence tests by Binet and Simon. They made a distinction between three different modes of assessing intelligence: 1) The medical method – which aims to find proof for an inferior intelligence by assessing anatomical, physiological and pathological signs; 2) The pedagogical method – focusing on making indirect assessments of intelligence by assessing the acquired knowledge; and 3) The psychological method – aiming to making direct assessment of intelligence with different tests of reasoning and memory (Ackerman, 1996). The psychological method measures aptitude or capacity of learning, which is also the aim in tests such as the SAT and SweSAT, while the pedagogical method measures achievement, such as different types of standardized achievement tests and grades.

The aptitude and achievement distinction is closely related to the distinction of Fluid intelligence (*Gf*), which represents reasoning and thinking in novel situations, and Crystallized intelligence (*Gc*), which represents culturally valued knowledge and skills achieved through schooling, made by Cattell (1963). Lohman's (2004) fluid-crystallized continuum is also based on the distinction between *Gf* and *Gc*, where tests measuring cognitive abilities can be found on the fluid end and school grades can be found closest to the crystallized end.

Predictive validity of *Gf* and *Gc*

According to the predictions from Investment theory, *Gf* should have its strongest influence on the development of knowledge and skills in younger ages and early school years (Cattell, 1987). There is research which investigates the predictive power of both *Gf* and *Gc* on early skills such as reading achievement, writing ability and mathematical achievement.

In a study by Evans, Floyd, McGrew, and Leforgee (2001), the relationship between the Cattell-Horn-Carroll (CHC) theory and reading achievement across childhood and adolescence, in a sample ranging between 6 and 19 years, was investigated. It was found that there was a strong relationship between *Gc* and reading achievement across ages. The relationship between *Gc* and reading achievement also seemed to increase up until the age of 15. However, no relation between *Gf* and reading achievement was identified. Francis, Fletcher, Maxwell and Satz (1989) found in a longitudinal study that there was an equally strong relation between *Gc* and *Gf* and reading achievement in grade 1. However, for reading achievement measured later, only *Gc* was important. De Jong and van der Leij (1999) found that there was a decrease in the relation between *Gf* and reading achievement between the first and the second grade.

Floyd, McGrew, and Evans (2008) investigated the relationship between the CHC ability measures and writing ability in a sample ranging between 7 and 18 years. They found that *Gc* was the strongest and most consistent predictor of writing ability across childhood and adolescence. McGrew and Knopik (1993) investigated the relation between CHC ability measures and writing ability in a wider age sample ranging between 5 and 79 years. Processing speed and *Gc* had the most consistent relation with writing ability. However, the relation with processing speed (the ability to perform simple

cognitive tasks fluently and quickly) decreased over the years while the relation with *Gc* increased. *Gf* had its strongest relation with basic writing skills in the early years (6-13 years) and had a consistent relation to written expression across all ages.

The relationship between CHC ability measures and mathematics achievement was investigated by Floyd, Evans and McGrew (2003) in a sample of children and adolescents between 6 and 19 years old. *Gc* had the strongest relation with both mathematical calculation skills and mathematical reasoning. After the age of 9 *Gc* only had a moderate relation with mathematical calculation skills, while the relation between *Gc* and mathematical reasoning was moderate but seemed to be strong after the age of 10 and for the remaining years. *Gf* had a moderate relation with mathematical calculation skills and a somewhat stronger relation with mathematical reasoning throughout childhood and adolescence.

The research above points to the importance of both *Gf* and *Gc* in predicting future achievement. While *Gf* seems to have a constant relation to certain skills and to be most important in the younger years, *Gc* seems to increase in importance throughout childhood and adolescence, which is much in concordance with the Investment theory.

The importance of *Gc* as a predictor of future achievement is also shown in adolescent and adult samples. In several studies *Gc* tends to display a superior predictive power when it comes to predicting achievement. Posthlethwaite (2011) showed that *Gc* was a superior predictor of several outcomes. In a meta-analysis of over 400 studies, Postlethwaite (2011) examined the ability of *Gf* and *Gc* to predict real world outcomes. It was shown that measures of *Gc* are stronger predictors of academic achievement, both at high school and college. Similarly, *Gc* was also found to be a better predictor of job performance than *Gf*.

Ackerman and colleagues have in a series of studies showed that *Gc* and prior knowledge are stronger predictors of future achievement than *Gf* in adult samples. In a study by Beier and Ackerman (2005) *Gf*, *Gc* and prior knowledge were investigated as predictors of learning new information about cardiovascular disease and xerography. A video presentation and a homework assignment were used as methods for learning. The sample consisted of 199 adults who were between 19 and 68 years old. The path analysis showed that *Gc* was a strong predictor of learning, both through a direct effect and through an indirect effect via prior knowledge. A direct effect from *Gf* was

identified on the video presentation, but there was no direct effect on learning from the homework assignment. Even though the effect of *Gf* was strong on learning it was indirect and mediated via *Gc*.

Ackerman, Bowen, Beier and Kanfer (2001) investigated different determinants of knowledge, including measures of *Gf* and *Gc* in Physical Science/Technology, Biology/Psychology, Humanities and Civics domains. The sample consisted of 320 university freshmen. Structural equation modeling was used to investigate the influence from the different determinants including *Gf* and *Gc*. *Gc* had substantial relations with all knowledge variables, while *Gf* only had a direct relation with the Physical Science/Technology variable.

The research above shows that *Gc* generally is a better predictor of achievement than *Gf* in adolescent and adult samples. This is also in accordance with Encapsulation theory, which predicts that *Gf* only will have direct relations with learning and achievement in younger samples and when there are measures of *Gc* available there will be no extra information provided by *Gf*. This is because *Gc* reflects the variance from *Gf* which is predictive of learning and achievement but also that it reflects individual differences in knowledge and skills which are of importance for future learning and achievement (Gustafsson & Carlstedt, 2006).

Predictive power of school grades

As indicated above, measures of *Gc* have a good predictive power when it comes to predicting future achievement. According to Lohman's (2004) continuum, school grades are placed closest to the crystallized end and thus reflect not only the information from *Gc* but also a wider array of knowledge and skills. Grades should be able to predict study success when they are used for selection and it is also important for educators that students have the knowledge and skills which are required to handle the education. Were it to be otherwise there would be consequences both for the individuals, but also for the quality of the education. It is also important that the selection is fair for the individuals, but also that it is effective in that the students who are best equipped to successfully complete the education are selected.

In most countries measures of previous school achievement or measures of capacity for learning are used when selecting students for further education. In Sweden students' grades alone are used for selection into upper secondary

education, while for selection into higher education both grades and the SweSAT, a test designed to measure capacity for learning, is used. The SweSAT is heavily loaded by *Gc* and aims to measure both verbal and non-verbal abilities as well as the capacity to handle information (Carlstedt & Gustafsson, 2005).

An abundance of research shows that measures of previous school achievement, including school grades, are better predictors of achievement than are measures of *Gc*, which in turn are better predictors than measures of *Gf* (e.g. Atkinson, 2004; Cliffordson, 2008; Gustafsson & Carlstedt, 2006). One possible explanation might be that school grades reflect all the information from *Gc* but also a wider range of knowledge and skills and, in addition, social-behavioral aspects which are of importance for future achievement (Gustafsson & Carlstedt, 2006). Beier and Ackerman (2005) also argue that when the criterion and predictor are better matched in terms of content and breadth, predictive power will increase. Measures of achievement and indeed grades are broader and more closely matched with curricular content than tests which measure cognitive ability or capacity for learning.

Atkinson (2004) argues that measures of achievement are better predictors of academic success since they have higher predictive power but also because they are more closely aligned to the curriculum and are less related to socioeconomic background (Atkinson, 2004; Atkinson & Geiser, 2009). Further, he argues that achievement tests which reflect curricular content would send a message to students that studying in school is the best way to prepare for further education (Atkinson 2004). Grades are also measures of achievement and they are similarly closely connected to the curriculum, and additionally reflect a wider representation of the curricular content.

Traditionally tests measuring capacity for learning have been seen as more rigorous measures of students' knowledge and skills (Geiser & Santelices, 2007), while grades have been seen as inadequate in communicating information about students' knowledge acquisition (Allen, 2005). This is because they are considered to reflect construct irrelevant variation, which distorts the interpretation of grades. There also are indications that grades are not comparable between schools and over time, and that they suffer from grade inflation (Gustafsson, Cliffordson & Erickson, 2014). However, despite arguments of the inappropriateness of grades, grades have been shown to be more powerful predictors of educational success than different forms of tests measuring capacity for learning and standardized achievement tests and,

consequently, are considered to be the most valid instruments for selection (e.g. Almlund, Duckworth, Heckman & Kautz, 2011; Bowen, Chingos & McPherson, 2009; Carroll, 1982; Cliffordson, 2008; Geiser & Santelices, 2007; Gustafsson, 2003; Gustafsson & Carlstedt, 2006).

The predictive pattern of school grades and standardized tests designed to measure capacity for learning, such as the SAT, is discussed in Almlund et. al.'s (2011) review of the power of personality traits as predictors of academic achievement, health, economic success and crime. They conclude that standardized tests like the SAT are good at predicting later achievement and occupational outcomes. However, in line with several other findings, they also acknowledge that cumulative GPA is a much stronger predictor of high school graduation and college-rank in class than standardized tests like the SAT. Almlund et al. (2011) argue that standardized tests and grades are associated with each other, but there are also obvious differences. While the SAT is meant to measure capacity for studies, grades measure achievement and are more closely aligned with the curriculum (Atkinson, 2004). Almlund et al. (2011) suggest that these differences in predictive pattern might be explained by personality factors such as conscientiousness. While standardized tests such as the SAT are more strongly related to *Gf*, school grades are more strongly related to different facets of conscientiousness, such as self-control and to conscientious behavior in the classroom. Thus, standardized tests measuring capacity for learning are good predictors of future achievement, but teacher assigned grades tend to be stronger predictors.

Most studies on the predictive validity of grades have investigated the predictive validity of high-school grades or upper secondary school grades in terms of admission to higher education. These studies have typically been conducted in an American context. Numerous studies have investigated the predictive power of different types of standardized admission tests, such as the SAT and high school grades, finding that grades are more accurate predictors of achievement (e.g. Camara & Echternacht, 2000; Fleming, 2002; Fleming & Garcia, 1998; Hoffman & Luwitzki, 2005; Komarraju, Ramsey & Rinella, 2012). Camara and Echternacht (2000) found in their literature review that high-school grades are better predictors of achievement than the SAT. This was also true when different criterion measures were used, such as freshman grades, graduation and cumulative GPA. Hoffman and Lowitzki (2005) also found that, for minority-group students, high-school grades were

stronger predictors of academic achievement than were standardized test scores.

In a Swedish context Cliffordson (2008) investigated the predictive validity of Swedish upper secondary school grades and the SweSAT in a sample of about 164 000 students born between 1974 and 1982. The sample included students who had graduated upper secondary school both with norm-referenced and criterion referenced grades. Two-level modelling was used to investigate the differences in predictive validity between upper secondary school GPA and the SweSAT. The results showed that upper secondary school grades were stronger predictors than the SweSAT. Despite the differences in design and purpose between the two grading systems they were equally strong in predicting success in higher studies. In fact, the criterion-referenced grades had somewhat stronger relation in predicting success in higher studies. The variability between programs and the impact of subject were also less noticeable for the school grades compared to the SweSAT.

However, school grades have also been shown to be related to and able to predict other outcomes such as job performance (Roth, BeVier, Switzer & Schippman, 1996). Miller (1998) investigated the effect of high school grades on later earnings. It was found that grades predicted long term earnings and productivity both for men and women even when controlling for background variables such as SES, region, ethnicity and school. Further, it was argued that some part of the gains in productivity can be explained by the soft skills that employers are asking for and that these also were reflected in grades.

A literature review on school, learning and mental health by Gustafsson, et al. (2010) discusses the relation between achievement and mental health. It was found that poor school grades and school failure are related to poor mental health. They also found that poor school grades have a relation with symptoms of depression, however this was only true for girls. There was also a relation between academic achievement and positive health outcomes.

6 Reflections on the theoretical framework

The previous studies and the theoretical perspectives described above give a backdrop to the predictive validity of different measures. There seems to be causality in the predictive pattern of cognitive abilities, running from *Gf* to *Gc*. This implies that, much in accordance with Cattell's (1987) Investment theory, there is a stronger influence of *Gf* in the early acquisition of knowledge and skills. However, even though learning is dependent on the level of *Gf*, the acquisition of knowledge and skills will increasingly rely on *Gc*. The increasing importance of crystallized abilities is also emphasized by Gustafsson and Carlstedt (2006), Ackerman (1996), Beier and Ackerman (2005), who explain that in adolescent and adult samples there is no predictive efficiency gained by adding measures of *Gf* when measures of *Gc* are available. The influence from *Gf* is then only indirect and mediated through *Gc*.

These results are also well in line with research that has found that measures of previous achievement are better predictors of academic success than are measures *Gc*, which in turn are better than measures of *Gf*. The predictive power of measures of achievement could be understood in the light of investment of intellectual abilities into the learning. While *Gf* has its strongest influence in early years of knowledge acquisition, Gustafsson and Carlstedt (2006) explain that the predictive power of *Gc* could be accounted for by such measures encapsulating all the variance from *Gf* which is predictive of learning and achievement, and also reflecting individual differences in knowledge and skills which are of importance for future achievement. Grades are located at the crystallized end according to Lohman's (2004) continuum, and thus they are also expected to reflect not only *Gf* and *Gc*, but also a broader array of knowledge and skills which are important for school achievement. The pattern of investment of different cognitive abilities can therefore help to explain the predictive power of measures of achievement and in particular of school grades.

Ackerman (1996, 2000) also emphasizes the importance of personality factors for the direction of the investment of intellectual abilities into knowledge acquisition. Several researchers have also indicated that grades potentially re-

flect such personality aspects, both indirectly but also directly through teachers' grading. The studies on the dimensionality of compulsory school grades by Klapp Lekholm and Cliffordson (2008, 2009) provide a basis for the investigation of whether and how the different dimensions, both cognitive and social-behavioral, contribute to the predictive power of school grades. By modelling the different dimensions in school grades it will be possible to better understand their predictive power.

7 Purpose

When grades are used for the purpose of selection it is important that they can also predict future study success, because there will be consequences both for individuals and for the quality of education if individuals do not have the necessary prerequisites for meeting the demands of the education. A selection system should also be fair and effective in selecting individuals who are most eligible for the education. Studies on the predictive power of school grades have indeed shown that they are able to predict study success, as well as a range of other outcomes (e.g. Atkinson Camara & Echternacht, 2000; Cliffordson, 2008; Fleming, 2002; Geiser & Santelices, 2007). However, why grades are such strong predictors of academic success is not quite clear. As indicated by theory and previous research, a partial answer could be found in the influence of both cognitive and social-behavioral aspects.

The purpose of the present thesis is to investigate the dimensionality of grades and the influence of both cognitive and social-behavioral dimensions for their predictive validity. The influence of cognitive aspects is investigated using Cattell's (1987) Investment theory, implying that the influence of *Gf* and *Gc* on the development of knowledge and skills throughout compulsory school is investigated. The predictive power of social-behavioral aspects is investigated by modeling the subject-specific and common-grade dimensions of school grades in accordance with previous studies conducted by Klapp Lekholm and Cliffordson (2008, 2009). They developed a model of grades with three subject-specific dimensions and one common-grade dimension. This model is used to investigate the dimensionality of both norm-referenced and criterion-referenced school grades and it is extended in order to investigate the predictive power of the different dimensions.

Investment theory and extensions of it developed by Gustafsson and Carlstedt (2006) and Ackerman (1996) can provide a tool for interpreting the influence of different cognitive dimensions on school grades, as well as the encapsulation of other aspects important for school achievement. Research on the dimensionality of grades and teachers' grading practices can further explain the pattern of dimensionality and predictive validity of grades.

More precisely the purpose is to investigate:

1. The stability of the dimensionality of criterion-referenced compulsory school grades across cohorts.
2. Similarities and differences in the pattern of dimensionality of norm-referenced grades compared to criterion-referenced grades.
3. The relative importance of the different dimensions in both criterion-referenced and norm-referenced compulsory school grades for their predictive validity.
4. The potential incremental effects of gender and parents' education on upper secondary school grades over and above the effects already mediated through compulsory school grades
5. The potential differences in patterns of dimensionality due to gender and parents' education.
6. The predictive pattern of cognitive abilities for the development of knowledge and skills, using Cattell's Investment Theory.

8 Method

A longitudinal design allows for the investigation of individual change over time, giving it certain advantages compared to a cross-sectional design. In order to investigate the relative influence of both cognitive and social-behavioral factors it was necessary to have data on the same individuals at different points in time. The availability of longitudinal data was taken advantage of in order to investigate the predictive pattern of cognitive and social-behavioral aspects.

Data

The data used within the studies in the present thesis was retrieved from the Gothenburg Educational Longitudinal Database (GOLD) and the Evaluation Through Follow-up Database (ETF). GOLD contains register data for all individuals in Sweden born between 1972 and 1992. Information on grades from compulsory school, national tests from grade 9, grades and study track from upper secondary school, gender and parents' education is available in the GOLD database, amongst other information. ETF is a longitudinal project which is built on 10 per cent nationally representative samples for 10 birth cohorts. In the ETF database information from standardized tests, knowledge tests, cognitive tests and grades in several different subjects are available.

In order to answer the questions of the stability of the empirical model created by Klapp Lekholm and Cliffordson (2008), and the relative importance of subject-specific dimensions and the common-grade dimension for the predictive validity of criterion-referenced compulsory school grades, population data from three consecutive birth cohorts born in 1987, 1988 and 1989 (each consisting of about 100 000 students) was used. Information on grades and national tests from compulsory school, grades from upper secondary school, gender and information on parents' education was used, all of which was retrieved from the GOLD database.

In order to answer the question of dimensionality and predictive validity of norm-referenced compulsory school grades, sample data on about 4000 students born in 1972 was used. Only students studying an academic track

preparatory for higher education in upper secondary school were included in the study since the grades in upper secondary school were not comparable between tracks. The data was retrieved both from the GOLD database and the ETF database. Information on grades and standardized tests from compulsory school, cognitive tests, grades from upper secondary school and information on gender and parents' education was used.

In the investigation of the influence on cognitive aspects on school grades, a 10 per cent nationally representative sample was used. The sample consisted of about 9000 students born in 1972. Data was retrieved from the ETF database and, since these individuals were followed-up in Grade 3, 6 and 9, it was possible to take advantage of these follow-ups in order to analyze the influence of *Gf* and *Gc* on the development of knowledge and skills. Information on grades and standardized tests in compulsory school, cognitive tests and a mathematical knowledge test was used.

Method of analysis

In the investigations of the dimensionality of grades and the influence of both cognitive and social-behavioral factors, structural equation modelling (SEM) was used. Structural equation modelling is closely related to multiple regression analysis (MRA), but enables the researcher to analyze both manifest and latent variables. Latent variables represent such constructs which are not directly observable. Through SEM it is possible to capture these underlying theoretical constructs and the relations between them. The advantage of latent variables is that they do not contain measurement error and they better can represent constructs which are formulated by the researcher through theory (as opposed to single manifest (observable) variables). Another difference to MRA is that SEM can handle several dependent variables and it is possible to have variables that are both dependent and independent and chains of variables. Hence, it is possible for the researcher to set up models which connect latent and manifest variables and investigate relations between these latent and manifest variables formulated by theory (Gustafsson, 2009).

Latent variables are representations of underlying constructs which are defined by a set of observed indicators. Relations between the latent variables and the indicators are set up in measurement models, where the latent variable is assumed to influence these observed indicators. Hence, through SEM it can be determined which indicators are influenced by the same underlying latent

construct. Based on theory and previous research, the researcher has to define the nature and number of factors, and which indicators that are related to which factors (Brown, 2006). Once it has been decided which indicators represent the construct, the relationships between constructs and their indicators are set up in measurement models. While the measurement model defines the number of factors and the relationship between the latent variable and its indicators, the structural model defines relationships between latent variables. The feasibility of models has to be discussed in relation to previous research and theory and trustworthiness, but the goodness-of-fit of the models is also tested against data.

It is important to bear in mind that the latent constructs are only representations or simplifications of real world phenomena, something that should always be borne in mind when interpreting results from latent variable modelling. However, in large scale studies one is often interested in the abstractions represented by the operationalization of theoretical constructs. The different dimensions investigated in the present thesis are representations of such underlying constructs; the subject-specific dimensions, the common-grade dimension and the intelligence dimensions (*Gc* and *Gf* being examples of these).

In the investigations of the dimensionality of both criterion-referenced and norm-referenced grades, measurement models were set up by subject grades and achievement tests from compulsory school as indicators of the subject-specific factors. The subject-specific factors account for the variation and co-variation among these indicators. The latent factor *ComGr* (the common-grade dimension) is related to all three subject grades. The *ComGr* factor can be perceived as an orthogonal factor aiming to measure variance in grades which is not present in the tests. In a structural model the subject-specific factors and the *ComGr* factor are then related to the grades in upper secondary school in order to investigate the predictive validity of the different dimensions (Figure 3).

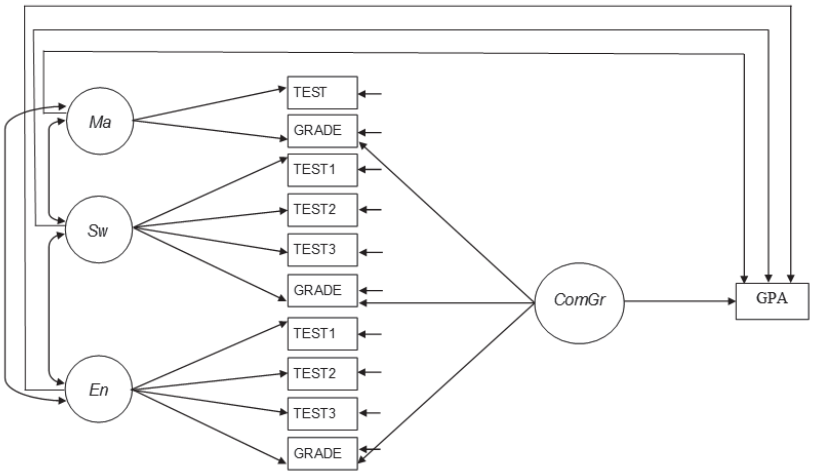

Figure 3. Model of subject grades, tests scores and the upper secondary school grades

Through SEM it was also possible to investigate the influence of *Gf* and *Gc* on the development of knowledge and skills. The factor representing *Gf* is set up by a metal folding test and a mathematics test, the factors of *Gc* in Grade 3 and 6 are set up by an antonym test and *Gc* in Grade 9 by three standardized tests. In an autoregressive model the relations between these factors are defined and regressed upon each other. (Figure 4).

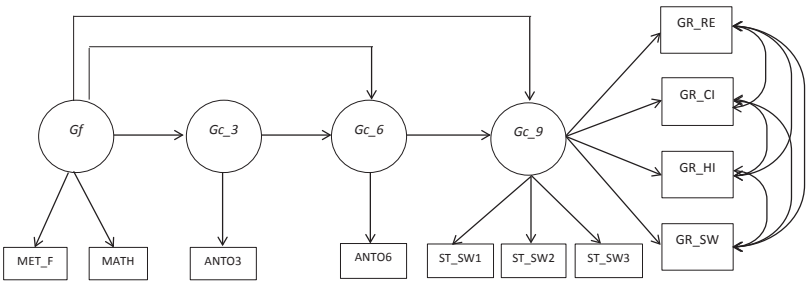

Figure 4. Example-model of the factors *Gf, Gc* and the subject grades in compulsory school

Goodness of fit

In SEM the models are evaluated on their feasibility and relation to theory, but they are also statistically tested against data. When the model is statistically tested the discrepancies between the sample variance-covariance matrix and the predicted variance-covariance matrix are determined, from which an

overall goodness of fit χ^2 test is computed. The χ^2 test is an established method for determining model fit but this test-statistic also tends to be inflated by sample size, which would result in rejection of virtually all models when the sample size is large. It is also criticized for being based on a very stringent assumption that the sample covariance matrix equals the predicted covariance matrix. Hence it is rarely used alone but rather together with other fit indices.

One such tool is the Root Mean Square Error of Approximation (RMSEA) which is relatively independent of sample and model size. As opposed to the χ^2 test which investigates if the model holds exactly in the population, the RMSEA tests if a model fits reasonably well in the population (Brown, 2006). Values below .08 indicate an acceptable model fit and values below .05 indicate a good model fit (Jöreskog, 1993).

The Standardized Root Mean Square Residual (SRMR) is an absolute fit index and evaluates how well a model fits the sample data. SRMR values can range from 0 to 1, however a good model fit is indicated by values below .05, i.e. the smaller the value the better model fit (Brown, 2006). However, values up to .08 are regarded as acceptable. The Comparative fit index (CFI) compares the χ^2 statistic of the model to the χ^2 of a more restricted null model. The CFI can range from 0 to 1, values closer to 1 indicating a good model fit (Brown, 2006).

Design effects

A common phenomenon when investigating issues in educational settings is that the data is clustered. Clustered data implies that students are situated in classes and schools, individuals belonging to the same cluster tending to be more alike than individuals belonging to different clusters (Gustafsson, 2009). These cluster effects or design effects (Muthén & Satorra, 1995) could be accounted for by using multilevel modelling. Multilevel modelling allows for explaining variance on several hierarchical levels, e.g. school, class and student (Heck, Thomas, Tabata, 2010). However, in the present context multilevel modelling would add to the complexity rather than to the understanding. Therefore cluster effects or design effects were handled by the "complex" option implemented in the Mplus program. Although this method compensates for clustering effects on the chi-square and standard errors it does not affect the estimates.

Missing data

Another common problem is that information sometimes is missing on different variables, which ultimately implies that information is lost on some variables (Schafer & Graham, 2002). In the studies included in this thesis, the data is assumed to be missing at random (MAR). When data is MAR it implies that missingness may depend on values that are observed (Allison, 2003; Schafer & Graham, 2002). MAR is a weaker assumption than the assumption that data is missing completely at random (MCAR). It is reasonable to assume that data is MAR in situations where missingness depends on observed data but not on missing data (Schafer & Graham, 2002).

Since the data contains missing information, the maximum likelihood missing data modeling procedure implemented in the Mplus program was used (Muthén, Kaplan & Hollis, 1987). This procedure uses available data to compute maximum likelihood estimates of model parameters, implying that it is the value that is most likely to have resulted in the observed values.

9 Results

The empirical part of this thesis consists of three articles, investigating different research questions. The studies involve the dimensionality of both criterion- and norm-referenced grades and the relative importance of the different dimensions in grades for their predictive validity. Study I and Study II also investigate the importance of the social-behavioral aspects for the predictive validity of grades, while the focus in Study III is the cognitive aspects of grades.

Study I

The main focus was on the importance of multidimensionality in criterion-referenced grades for their predictive validity and on the stability of the dimensionality across cohorts. It has been demonstrated in previous research that grades are superior in predicting future study success (e.g., Atkinson & Geiser, 2009; Carroll, 1982; Cliffordson, 2008; Gustafsson & Carlstedt, 2006). The question why grades display this pattern of predictive validity is not entirely clear. It was hypothesized that the different dimensions previously identified in grades might contribute to explaining the predictive power of grades. By the use of SEM, the relative importance of the subject-specific dimensions and the common-grade dimension for predicting educational success in upper secondary school was investigated. These issues were investigated on criterion-referenced grades for three consecutive birth cohorts. The common-grade dimension was also investigated with respect to stability over cohorts and predictive validity. Another purpose was to investigate effects of gender and educational background.

Three subject-specific dimensions and a common-grade dimension were identified. The common-grade dimension cuts across both grades and teachers, indicating that the dimension reflects aspects considered important by teachers across subjects. The results were much in line with previous findings of dimensionality by Klapp Lekholm and Cliffordson (2008). The pattern of results was similar in all three birth cohorts, demonstrating the stability and strength of the model and indeed the common-grade dimension.

Both the subject-specific dimensions and the common-grade dimension contributed to prediction of study success in upper secondary school as measured by course grades in English, Swedish and mathematics and the GPA, with a pattern similar for all three birth cohorts. The interpretation of the common-grade dimension was not entirely clear, but since it cuts across grades and teachers it was hypothesized to represent aspects which are valued by many teachers, explaining variance in grades over and above the subject-specific knowledge and skills.

It could be argued that the effect of the common-grade dimension is quite weak and that its importance should not be overestimated. Getting significant results of small effects is a clear danger when one has access to large samples. However, it is reasonable to believe that if it would have been possible to use all 16 grades from compulsory school there should be an equal contribution, or higher, from these subject grades. Hence, even though the effect can seem small, it would be significantly larger in the light of the contribution of several subject grades.

Questions of gender differences and differences due to parents' education also were investigated. It was found that apart from the influence that gender had on compulsory school grades there were small incremental effects on the upper secondary grades, apart from Swedish where girls were additionally favored, and English where boys were additionally favored. Furthermore, the results showed that girls had a significant advantage in the common-grade dimension.

There were also differences in the common-grade dimension due to the students' educational background. Students with lower educational background were favored in the common-grade dimension which is in concord with the results found by Klapp Lekholm and Cliffordson (2009). They interpreted these differences as being the result of a compensatory grading practice.

Study II

An identical model used to investigate criterion-referenced grades was used to investigate the dimensionality of norm-referenced grades. The predictive power has in previous research been shown to pertain to both norm-referenced and criterion-referenced grades (Cliffordson, 2008). However, there are reasons to believe that the norm-referenced grades would not display

the same pattern of multidimensionality as the criterion-referenced grades since the standardized tests used as a support for grading within the norm-referenced grading system were more influential than the national tests used within the criterion-referenced system. If the standardized tests had a stronger influence on teachers' grading it is possible that there would be a lesser influence of social-behavioral aspects in the grading process.

The purpose was to investigate the multidimensionality of norm-referenced compulsory school grades and the relative importance of the different dimensions for their predictive validity. In addition effects of gender and parents' education were investigated. Norm-referenced compulsory school grades in mathematics, English and Swedish and standardized achievement tests in corresponding subjects were used in order to investigate the dimensionality of norm-referenced compulsory school grades. Only students studying a program preparatory for university were included since the courses in upper secondary school were not comparable between vocational and academic tracks

The main results showed that norm-referenced grades were multidimensional, reflecting both subject-specific knowledge and skills, and a common-grade dimension. The common-grade dimension found in norm-referenced grades accounted for a somewhat larger part of the variance in norm-referenced grades than was found in criterion-referenced grades.

When the different dimensions were related to upper secondary school grades in mathematics, English and Swedish and students' GPA scores, both the subject-specific dimensions and the common-grade dimension contributed to predict upper secondary school grades. The common-grade dimension identified in norm-referenced grades had a somewhat stronger relation to upper secondary school GPA than the common-grade dimension identified in criterion-referenced grades. This stronger relation seems paradoxical in the light of the higher regulating role of the standardized tests. However, the standardized tests only regulated the grading on the class level, not the individual level. Thus, the teacher had the freedom of ranking students within the class, which implies that there were opportunities to consider social-behavioral aspects in the grading process.

When the effects of gender and parents' education were investigated, small incremental effects were found on the subject grades and GPA in upper secondary school, in addition to that already mediated through the compulsory school grades. The results also indicated that girls and students with higher

educational background were favored in the common-grade dimension. This result runs contrary to the result found on criterion-referenced grades where students with lower educational background were favored in the common-grade dimension. These results can most certainly be explained by the restricted sample composed of students enrolled on academically-oriented tracks. This implies that, irrespective of background, these students have pressure and support at home to succeed in their studies. Thus the differing findings in the studies cannot be directly compared since the composition of students in the two studies differs.

Study III

While Study I and II focused primarily on the social-behavioral aspects of the dimensionality of grades, Study III aimed to investigate the cognitive aspects of grades. In particular, Study III aimed at investigating the influence of *Gf* and *Gc* on the development of knowledge and skills by the use of Cattell's Investment theory (1971, 1987), thereby also testing this theory. Previous studies have found that the importance of *Gf* in predicting future achievement diminishes when individuals grow older and when measures of *Gc* are available (Ackerman, 1996; Gustafsson & Carlstedt, 2006). Thus, it was expected that *Gf* would not contribute with additional information when measures of *Gc* were available. Rather, it was expected that results would be in line with Encapsulation theory stating that all the information from *Gf* is encapsulated in *Gc*, and all the information from *Gc* is encapsulated in grades (Gustafsson & Carlstedt, 2006).

By the use of a cognitive metal-folding test in grade 3 and a mathematical knowledge test, a latent variable representing *Gf* was set up. *Gc* was in grade 3 and 6 measured by a cognitive antonym-test, and in grade 9 by standardized tests in Swedish and English. A model was formulated where *Gf* was related to *Gc* in Grade 3, which in turn was related to *Gc* in Grade 6, and in turn again related to *Gc* in Grade 9, which, finally, was related to grades in Swedish, History, Civics and Religion in Grade 9. A second similar model was also formulated where the standardized test in English in Grade 9 was used to define *Gc* in Grade 9 and the final outcome measure was the subject grade in English.

Even though it was expected that *Gf* would not contribute with additional information when measures of *Gc* were available, such a model marginally

achieved an acceptable model fit, although there was room for improvement. When direct relations from *Gf* to *Gc* in both grade 6 and 9 were introduced, in line with Investment theory assumptions, the model fit improved. These results indicate that there is a continuous influence from *Gf* on the development of knowledge and skills in line with Investment theory assumptions throughout compulsory school. The results show that the strongest influence from *Gf* occurs in grade 3, but there is also a continuous influence from *Gf* in grade 3 and 6, even though these effects are much weaker. One hypothesis to account for this is that when children go through the school system they meet tasks of increasing complexity, thus demanding an increasing investment of *Gf*. It is also possible that these results might be due to the fact that the students in the present context are considerably younger than for example in Gustafsson's and Carlstedt's (2006) study, implying that a significant investment of *Gf* is needed in the early school years. Furthermore, there were considerable indirect effects of *Gf* on *Gc* and grades throughout compulsory school. These results support Cattell's (1971, 1987) theory that *Gf* is the driving force in knowledge acquisition in school and that *Gf* type abilities influence the development of knowledge and skills both directly and indirectly through *Gc*. However, no direct effects of *Gf* on the subject grades could be observed, giving partial support to Encapsulation theory (Gustafsson & Carlstedt, 2006).

Previous studies of Investment theory have shown somewhat inconsistent results, thus pointing to the complexity of the theory. However, there are studies which have shown the importance of sample homogeneity in investigating Investment theory (e.g. Kan et al., 2011; Valentin Kvist & Gustafsson, 2008). Valentin Kvist and Gustafsson (2008) found that, in homogenous samples, *g* and *Gf* have a correlation of unity and these findings were extended by Kan et al. (2011) showing that *Gc* equals verbal comprehension. Thus, following the argument by Valentin Kvist and Gustafsson (2008) that if different subgroups have had different opportunities to learn, the simple relationship between *Gf* and *g* will break down, it was argued that the sample homogeneity may be playing a role in the direct relations from *Gf* to *Gc* in grade 6 and 9 which were observed.

10 Discussion and Conclusions

Dimensionality of grades

The results show that both criterion-referenced and norm-referenced compulsory school grades are multidimensional, reflecting both subject-specific dimensions and a common-grade dimension. This provided an answer to Research Question 2, in showing the similarities in dimensionality in the different grading systems. When investigating Research Question 1 the results also showed that the dimensionality of the criterion-referenced grades was stable across several cohorts. The multidimensionality of the criterion-referenced and the norm-referenced grades contributed to explain their predictive validity, which provided an answer to Research Question 3. The subject-specific dimensions accounted for the major part of the variance in grades, indicating that the teachers' grading is in substantial correspondence with the national tests, which corresponds with similar findings in studies carried out by the National Agency for Education. This also was true for the study on norm-referenced grades, although the correspondence was somewhat smaller. The common-grade dimension also accounted for variance in grades. Both the subject-specific and the common-grade dimension contributed in explaining the predictive validity of grades.

The common-grade dimension identified in both grading systems cuts across both subject grades and teachers, implying that it reflects aspects common to all three subject grades and all teachers. These commonalities, considered by many teachers, is a result of a direct influence on grades from teachers' grading. However, it is harder to discern whether this direct influence is causing an irrelevant variance in grades or whether it is caused by aspects which are relevant to the construct of grades. It is however reasonable to believe that the influence is both relevant in relation to the construct of grades and, to some extent, also irrelevant.

The common-grade dimension

As discussed above, both norm-referenced and criterion-referenced grades displayed a pattern of dimensionality, with a common-grade dimension, which

is in line with the research on criterion-referenced grades by Klapp Lekholm and Cliffordson (2008, 2009). In further investigation of the common-grade dimension Klapp Lekholm and Cliffordson (2009) found that it was related to students' general interest in studies and to parents' commitment in their children's learning. Studies on the discrepancies between grades and standardized achievement tests have related these discrepancies to represent social-behavioral aspects (e.g. Bowers, 2008, 2011). Hence, previous research on grades has directly or indirectly shown that there is variance in grades which could be related to aspects which are not a part of the knowledge and skills stated in the curriculum. The present study supports these interpretations, suggesting that the common-grade dimension is, in part, representing social-behavioral aspects. Additionally, as indicated in previous research (e.g. Brookhart, 2013; Klapp Lekholm & Cliffordson, 2009), the current findings suggest that common-grade dimension is related to teachers' grading practices.

It could also be argued that the achievement tests in compulsory school used as measures of subject-specific knowledge and skills are not perfect measures and, to some extent, reflect social-behavioral aspects. However, even though the achievement tests are by no means perfect measures, the discrepancies defined as the common-grade dimension indicate that there are aspects which pertain to all three subjects and which influence grades. Additionally, if the tests were perfectly objective measures of students' knowledge and skills, the common-grade dimension would arguably have higher values.

Commonalities in the syllabi

Whether this influence on grades is irrelevant to the construct of grades or is an expression of commonalities in the different syllabi for the different subjects is hard to discern. When comparing the different syllabi for English, mathematics and Swedish in Lpo94, there are commonalities which primarily pertain to the goals which every student should strive for. These commonalities are primarily related to verbal abilities, such as being able to both orally and in written production, communicate and argue for one's thoughts and to be able to express oneself clearly. Hence, it is possible that the common-grade dimension is, in part, a reflection of the verbal abilities expressed in all three syllabi. The influence of verbal abilities is also indicated in the research by Gustafsson and Balke (1993). In their investigation of the

general school achievement factor the influence of verbal abilities was substantial. However, when comparing the different syllabi for the different subjects for the norm-referenced grades, these commonalities are not as evident. For Swedish and English there are clear indications that verbal abilities are important for both subjects, but this is not evident in the syllabi for mathematics. Yet, it is reasonable to believe that the verbal abilities are also important for the norm-referenced mathematics grade, and certainly for being able to communicate ideas and thoughts.

It is also possible that the common-grade dimension reflects aspects which are important for all school subjects and which are also expressed as overall goals in the curriculum. These could be aspects such as being able to cooperate, taking responsibility of your own studies and showing respect towards others. Still, these are aspects which are not intended to be a part of the grading.

Social-behavioral aspects

As indicated in previous research, the common-grade dimension might also be related to an influence of social-behavioral aspects which are irrelevant to the construct of grades, but which are considered by teachers in their grading. These aspects could be many and apart from interest and parents' commitment in their children's studies, other studies have identified that teachers tend to consider a plethora of aspects, such as attitude, effort and engagement (e.g. Brookhart, 1991, 1993; Randall & Englehard, 2010), and that achievement and effort is confounded (Brookhart, 1993). Since the common-grade dimension is related to all grades and teachers there seem to be aspects which are considered important by many teachers. It is reasonable to believe that these are aspects which are related to the management of the classroom, such as being able to cooperate, taking and following instructions, and being able to work autonomously. McMillan (2001) also shows that the management of the classroom is important especially in encouraging motivation and student engagement and understanding, and that teachers inevitably make assessments that motivate students and encourage learning (McMillan, 2003). The motivating function is also an important aspect of grades, which is also stated as a curricular goal. It is however possible that such assessment practices may lead to a confounding of both knowledge, skills and social-behavioral aspects.

Another line of research on halo effects might also provide a possible layer of explanation to the meaning of the common-grade dimension. Even though research on halo effects in educational settings is sparse, there is research which suggests that in teachers' ratings there is a substantial influence of halo effects (e.g. Dennis, 2007). It is reasonable to believe that the overall impression of a student will affect the assessment of aspects and attributes which are not entirely known to the teacher.

Hence, there are several possible interpretations of which aspects are reflected by the common-grade dimension. Although the common-grade dimension has previously been found to be related to general interest in school work and parents' commitment (e.g. Klapp Lekholm & Cliffordson, 2009), research on teachers' grade assignment has found that other aspects which are related to the motivating function of grades, management of the classroom and/or of overall impressions of students are also possible interpretations (e.g. Brookhart, 1993; McMillan, 2001, 2003).

It is also possible that there is an indirect influence of social-behavioral aspects on grades. However, such indirect influence would, in the present study, result rather in stronger subject-specific dimensions. One example of this could be that when students see that they get rewarded for putting effort into the school, motivation increases, which could be reflected in subsequent achievement, which would result in an indirect influence.

Validity issues

The results of dimensionality discussed above indicate that there could be threats to the validity of grades. The identification of the common-grade dimension indicates that grades could be suffering from construct irrelevant variance, implying that teachers consider aspects which are not part of the construct of grades (Messick, 1989), something also indicated in previous research (e.g. Klapp Lekholm & Cliffordson, 2009). As discussed above it is however also reasonable to believe that part of the variance in the common-grade dimension could be a result of commonalities in the syllabi for different subjects. Still, the results here indicate that there are threats to the validity of grades and that they suffer from construct irrelevant variation and construct underrepresentation.

Construct irrelevant variance and underrepresentation

The correspondence between the achievement tests in compulsory school and grades is quite high, which indicates that teachers manage to capture the aspects tested in the achievement tests very well. This result is in line with previous research on the correspondence between standardized achievement tests and grades (e.g. Harlen, 2005; Johansson, Myrberg & Rosén, 2012; Südkamp, Kaiser and Möller, 2012). Still, these studies also show that there is a large variability between teachers which indicates that there could be threats to the validity of grades. However, neither tests nor grades are perfect measures of achievement. On the one hand, tests do not capture all of the goals in the curriculum and are only measures of achievement captured on a single occasion. On the other hand, grades are multidimensional as stated above, and the awarding of grades to some extent always implies a certain amount of subjectivity. A high correspondence between tests and grades will not necessarily imply more valid grades; it could rather lead to construct underrepresentation since the achievement tests in compulsory school do not test all aspects of the curriculum.

Influence of irrelevant aspects in grades has also been indicated elsewhere (e.g. Brookhart, 1991, 1993; Cizek et. al., 1996; McMillan, 2001; Randall & Englehard, 2010). Brookhart (1991) argues that this 'hodgepodge' of grading practice is due to teachers being aware of the effects of grades. In order to compensate for the consequences of grades, teachers tend to base grading on both knowledge and skills, and social-behavioral aspects. This type of grading practice can also be seen as an aspect of fairness and an important aspect related to the motivational consequences of grades (Brookhart, 1993). Hence, the inclusion of irrelevant factors in grading is a complex issue where teachers are concerned with both fairness and the consequences of grades. Social-behavioral aspects, such as being able to cooperate, are also emphasized in the curriculum as overarching goals, indicating that they are important goals of schooling even if these are not intended to be a part of subject grades.

However, even though this type of grading is carried out in line with the argument that it supports fairness, or that these also are important goals of schooling, there is a risk that it will eventually cause the construct of grades to be distorted, the construct validity of grades called into question (Allen, 2005). Grades which reflect construct irrelevant variance would ultimately lead to se-

vere consequences when they are used for making decisions regarding students or when they are used for selection purposes (Messick, 1989).

Social consequences

Even though there are indications that the direct influence of social-behavioral aspects contributes to explaining the predictive power of grades, such influence would not be without consequences when grades are used for purposes of selection (Messick, 1989).

In Sweden great trust is placed in the professional competence of teachers and teachers have substantial freedom in choosing content, assessment formats, carrying out their own observations of students and also valuing their own interpretations and validating their own inferences. Without any external validation there are several possible threats to validity in the form of irrelevant variance and underrepresentation of constructs in the grading practice.

The national tests distributed to all schools are to function as tools for equity in grading and for concretizing the goals in the curriculum. However, in addition to the fact that there are discrepancies between the tests and grades, there is also a great variability between teachers. There might additionally be problems with the scoring of the national tests, as revealed by the School Inspectorate (2012). This is an indication that there could be threats to the construct validity of grades on several levels and their usefulness for selection could be negatively affected. The purpose of selecting the individuals who are best equipped for the education might be jeopardized, which is important not only for individuals, but also for the quality of education (Wikström & Wikström, 2012). The results here, indicating a direct influence of social-behavioral aspects, suggest that grading practices are lacking in transparency and that there are implicit and tacit grounds for assigning grades (Klapp, 2010), meaning that, ultimately, the equality principles of schooling can be placed in jeopardy.

Still, it is important to note that the correspondence between tests and grades as indicated by for example Johansson (2012) is fairly high. There are also commonalities in the syllabi for the different subjects which relates to verbal ability, which would indicate that not all the direct influence of teachers' grading is a result of irrelevant aspects. There are also indications that the mean values of the national tests differ between different years, indicating that there also are problems with the comparability of the national

tests (Gustafsson et. al. 2014). Additionally, as argued by Messick (1989), it is necessary to compare one type of assessment with other types of assessment with the same purpose. When comparing grades to other measures of achievement or measures of capacity for learning, grades still have a higher predictive power, as well as capturing a wider array of knowledge and skills and, more importantly, grades constitute a better representation of the curricular content. In addition the reflection of social-behavioral aspects is revealed in both the study on criterion-referenced grades and in the study on norm-referenced grades, to be relevant and important for further study success. Levin (2011) also emphasizes the increasing importance of social-behavioral aspects as a societal aim and argues that these should be included in both small and large scale assessments.

Predictive validity

The results in the present thesis suggest that the predictive power of compulsory school grades is possible to explain with reference to the influence of both cognitive and social-behavioral aspects.

Gf and *Gc*

When investigating Research Question 6, the results indicate that fluid abilities have influence on the development of knowledge and skills throughout compulsory school. These predictions are in accordance with Cattell's (1971, 1987) Investment theory. However, the results are not fully in accordance with previous studies on the influence of *Gf* and *Gc* on the development of knowledge and skills. Rather, previous studies have typically found that the importance of *Gf* decreases as a function of age and that *Gf* does not provide any extra predictive power when there are measures of *Gc* available (e.g. Ackerman, 1996; 2000; Gustafsson & Carlstedt, 2006. However, these studies have had adolescent and adult samples, which could in part explain these differences.

The results in the present thesis suggest that the assumption of Investment theory that *Gf* is continuously invested into the development of knowledge and skills is supported. *Gf* influences the development of knowledge and skills throughout compulsory school and the direct effects of *Gf* on the measures of *Gc* were observed from grade 3 to grade 9. The effect from *Gf* was strongest in the third grade but, contrary to what was expected, there was also an effect

of *Gf* in grade 6 and grade 9, even though these were considerably weaker. There were also considerable indirect effects of *Gf* on subject grades, indicating that *Gf* type abilities – such as being able to induce, deduce, solve problems and conduct abstract reasoning – are important aspects of knowledge acquisition, both directly and indirectly through *Gc*, a finding which is in line with Cattell's (1987) Investment theory. The importance of *Gf* throughout compulsory school might be due to the fact that when students go through the school system they meet tasks of increasing complexity, which will demand an investment of *Gf*.

The continuous influence of *Gf* could, in contrast to previous studies, possibly be explained by the homogenous sample. The importance of homogeneity in investigating the assumptions of Investment theory have been demonstrated by both Kan et al. (2011) and Valentin Kvist & Gustafsson (2008). The larger predictive validities found for *Gc* measures in previous research can also be explained by age differences between the individuals included in the studies, implying that age is a factor which affects the results. The participants in Ackerman's, and Gustafsson and Carlstedt's studies were adolescents or adults, while the participants in the present study were in their early school years. In the present study *Gc* also seems to have a stronger influence towards the end of compulsory school when individuals have reached adolescence. Even though the results showed that there were considerable indirect effects of *Gf* on the measures of *Gc* and on the subject grades in Grade 9, there were no direct effects of *Gf* on the subject grades. When the broad measures in the end of compulsory school are available, *Gf* does not add any predictive force; instead the crystallized abilities seem more important for further knowledge acquisition.

It might also be that the continuous influence of *Gf* can be explained by the narrower definition of *Gc* in the present study than the definition used by Gustafsson and Carlstedt (2006), which includes mathematical ability. Gustafsson and Carlstedt (2006) also argue that broader measures of *Gc* are expected to have a stronger predictive power than narrower *Gc* measures including only verbal ability. This proposal can also help to explain why measures of previous achievement have a stronger predictive power than measures of both *Gf* and *Gc*. Measures of *Gc* reflect the information from *Gf* which is predictive of learning, but also such knowledge and skills which are important for future achievement. Measures of achievement, and indeed also

grades, reflect all the information from *Gc*, but also a wider array of knowledge and skills important for school achievement.

Measures of achievement

As indicated by the results presented here, as well as by previous research, the broadness of the measure plays a role in predicting future achievement. Not only is the wider array of knowledge and skills represented in measures of achievement important, but the representation of curricular content also seems to play an important role (Ackerman, 2004). The results in the present thesis indicate that the encapsulation of both *Gf* and *Gc* and the wider aspects of knowledge and skills which are closely connected to curricular content can, in part, explain the predictive power of school grades.

Accordingly, the compulsory school grades in Swedish, English and mathematics have a stronger power in predicting upper secondary school grades than results on respective achievement test in compulsory school. It might be argued that this is due to the fact that the subject grades reflect an even broader scope of knowledge and skills than the achievement tests. The identification of a common-grade dimension indicates that there are also aspects which are common to all subject grades, such as verbal ability and social-behavioral aspects, which are encapsulated in grades and which contribute to explaining why grades have such high predictive validity. The results also show that it is not only cognitive aspects that play a part in explaining the predictive validity of school grades. The predictive validity of grades could be explained by the fact that grades reflect a continuous influence of *Gf* and *Gc* and, in addition, that grades reflect a wider range of knowledge and skills as well as social-behavioral aspects which are predictive of success in further education (Gustafsson & Carlstedt, 2006; Thorsen & Cliffordson, 2012; Thorsen, 2014).

The importance of personality aspects has also been recognized by Cattell (1987) and even greater emphasis on such aspects is put forward by Ackerman (1996), who argues that personality and interest factors determine the direction and intensity of intellectual investment, and are thus reflected in measures of *Gc*. Gustafsson and Carlstedt (2006) also discuss the possibility that social-behavioral aspects are important in knowledge acquisition and are reflected in grades. It is important to note that the investment of such social-behavioral skills which both Ackerman (1996, 2000) and Gustafsson and

Carlstedt (2006) discuss primarily refers to the indirect influence of social-behavioral aspects which result in increased learning. The common-grade dimension identified in the present study rather signifies a direct influence of social-behavioral or verbal aspects, which is related to teachers' grading practice. These direct effects of aspects considered by teachers in their grading practice also have importance for future achievement, which is indicated by the contribution which the common-grade dimension makes to predictive validity.

All in all, the results in the present research indicate that the investment of both cognitive and social-behavioral aspects can explain the predictive validity of grades. Whether these aspects influence grades directly in terms of being aspects teachers consciously consider, or indirectly through increased learning, they are important for the prediction of subsequent achievement. Apart from the influence of *Gc* on school grades, there is also a substantial influence of *Gf*, which contributes to explaining the predictive validity of grades. It is also reasonable to believe, as discussed above, that the broadness of grades and the closeness to curricular content also contribute in explaining the predictive power of school grades. Another reason why grades are such good predictors of future achievement is that they are based on displayed achievement in many different forms over a long time, and the GPA summary is based on assessments by several different teachers.

Differing results due to grading system

When investigating Research Question 2, it was found that the two grading systems are similar in patterns of multidimensionality. However there were also differences concerning the common-grade dimension. The common-grade dimension identified in norm-referenced grades had a stronger relation with upper secondary GPA than the common-grade dimension identified in criterion-referenced grades. This can possibly be explained by differences in what the common-grade dimensions in the two grading systems represent. The common-grade dimension in norm-referenced grades is a better representation of social-behavioral aspects, while the common-grade dimension in criterion-referenced grades is, to a larger extent, a representation of a compensatory mechanism.

Because of the construction of the criterion-referenced grades with a "pass limit", and restrictions on admittance to upper secondary school and higher

education, it is possible that teachers favour lower-achieving students in that they take consideration of the social consequences a failing grade would have for the student. The common-grade dimension in criterion-referenced grades might, to a greater extent, represent a compensatory grading practice, where students are rewarded for effort or interest and awarded a passing grade even though they do not reach the goals in the curriculum. The differences found in the influence of parents' education between the two studies might also support this conclusion; that is to say that in the study on criterion-referenced grades, students with lower educational background had an advantage in the common-grade dimension, indicating a compensatory mechanism, while in the study on norm-referenced grades the results were the opposite. Furthermore, such a watershed or passing limit was not present to the same extent in the norm-referenced grades, since a student would not be prohibited from applying to upper secondary school if he or she received a grade 1 or 2.

This could also indicate that the common-grade dimension in norm-referenced grades and criterion-referenced grades differs with respect to the influence of social-behavioral aspects. The common-grade dimension in norm-referenced grades could be a better representation of social-behavioral aspects since the standardized tests only regulated grading on class-level, and since there is not as strong an influence of compensatory mechanisms. This could also explain why the common-grade dimension in norm-referenced grades is stronger in predicting subsequent achievement. However, due to limitations in the study on norm-referenced grades, this should be interpreted with caution since the sample was limited to students studying an academically oriented track preparatory for higher education.

It is also possible that the presence of compensatory grading would be even more actualized with the implementation of the new curriculum Lgr11. Students now need to pass a larger number of subjects to get into upper secondary education. It is possible that this would lead to an increase in the compensatory grading practice, since teachers seem to be taking account of the consequences that grades have for students.

Effects of gender and parents' education

Another purpose of the present thesis was to investigate whether there were effects of gender and parents' education in the common-grade dimension of grades, and whether there were incremental effects of these factors in the

upper secondary grades over and above what is already mediated through the compulsory school grades.

Gender

When investigating Research Question 4, the results indicate that there were small incremental effects on the criterion-referenced upper secondary school grades over and above the gender differences already mediated through the compulsory school grades. However, there were no incremental gender effects on the norm-referenced grades. In the investigation of Research Question 5, the results also showed that girls' values are higher in the common-grade dimension of grades, both in criterion-referenced and norm-referenced grades. Girls were also found to be favored in the common-grade dimension in the study by Klapp Lekholm and Cliffordson (2008, 2009), where it was found that almost all of the gender difference in the common-grade dimension was explained by interest for schoolwork. The advantage for girls in the common-grade dimension may partially be explained as a direct effect of interest; meaning that they show more interest and motivation in their schoolwork and also get extra credit for this by the teacher. However, girls' higher grades could also partially be explained by an indirect influence of interest, implying that they are more motivated for and interested in school work which results in increased learning and thus higher grades.

Other studies have suggested that girls and boys display somewhat different patterns of attitudes to school and school work, implying that girls are more motivated for school work and have attitudes which are favored in school (Dwyer & Johnson, 1997). It is possible that such attitudes, apart from having an influence on achievement, are also favored by teachers and might to some extent be considered in their grading, which is also indicated in girls' advantage in the common-grade dimension. However, all of the difference in grades between boys and girls cannot be explained by girls being given extra credit for being motivated and responsible. The National Agency for Education (2006) found that some of the differences in grades between boys and girls can be explained by the fact that girls seem to put more time into and have greater engagement in their school work. Time investment and engagement might arguably lead to higher achievement, even if it is possible that they also get extra credit for this when being graded.

There is certainly no single factor which can explain gender differences in grades, and there are several different directions of theories in explaining these differences, such as that the girls' grades now being more in line with their results on cognitive tests (Wernersson, 2010), or that girls might be compensated for their supposedly subordinate position in the classroom and thus be rewarded for being helpful. The interpretation in the present study is that girls' relatively higher grades seem to be a consequence of both indirect and direct effects; indirect in the sense that social-behavioral aspects influence learning per se which leads to higher achievement, and direct by teachers considering social-behavioral aspects when assigning grades, and which are thus reflected in the common-grade dimension.

Parents' education

When in Research Question 4 investigating the incremental effects of parents' education on the upper secondary school grades, the results indicated that such effects were evident in both grading systems. When investigating Research Question 5, effects of parents' education in the common-grade dimension of grades were found in the studies of both criterion-referenced and norm-referenced grades. However, these effects were completely reversed between the studies. In the study on criterion-referenced grades students from homes with lower education had higher values in the common-grade dimension, while students with higher educational background had higher values in the common-grade dimension in the study of norm-referenced grades. These differences could have several explanations. One interpretation might be found in the compensatory mechanism in the criterion-referenced system (Klapp Lekholm & Cliffordson, 2009). This implies that students who are disadvantaged might be compensated for their disadvantaged situation by teachers assigning a greater value to social-behavioral aspects.

The advantage for students from higher home educational backgrounds found in norm-referenced grades could rather be explained by the restriction of the sample to students only studying a track preparatory for higher education. Such students are highly likely to receive parental support and students with these backgrounds might also be better prepared for school life and more aware of the social aspects which are reflected in grades.

Limitations and further research

One limitation affecting the research carried out is the assumption that achievement tests in compulsory school only measure subject-specific knowledge and skills and it is certainly possible that achievement tests also reflect social-behavioral aspects. However, as discussed above, if the tests were pure measures of subject-specific knowledge and skills, this would result in an even larger common-grade dimension than that identified in the present research.

Another point is the contribution of other subject-grades to the common-grade dimension, which would have been interesting to investigate. It is reasonable to believe that the contribution to the common-grade dimension would be equally large or larger from other subjects. Because the achievement tests tend to regulate the influence of social-behavioral aspects in the grading process, in subjects without such tests the influence of social-behavioral aspects could be even greater. However, a prerequisite for investigating these issues is the availability of achievement tests, which, regrettably, is not the case with other subjects. However, with the implementation of Lgr11 national tests have been implemented in more subjects, so this would be an interesting line of future research.

Another limitation might be that, even though the use of dummy variables is a simple and effective method for investigating additive effects of gender and parental education, it is not possible to investigate possible violations to measurement invariance. In order to investigate such possible violations it would be necessary to conduct a multi-group analysis by splitting the sample into different groups. With such an approach it would also be possible to investigate, for example, if there were differences with respect to the predictive validity of compulsory school grades for different tracks in upper secondary school, i.e. between vocational and academic tracks.

Finally, the study would have benefitted from having information about social-behavioral aspects which could have influenced the common-grade dimension. Influence of such aspects on the common-grade dimension has already been investigated by Klapp Lekholm and Cliffordson (2009). Moreover, the indications of the importance of measures of personality for the investment individuals put into knowledge acquisition suggest that it would in future research be interesting to include measures of interest and personality

in order to investigate the nature of the interplay of personality factors in intellectual investment over time.

Conclusions

Grades are multidimensional, reflecting both subject-specific dimensions and a common-grade dimension. The subject-specific dimensions reflect knowledge and skills as defined by the achievement tests in compulsory school. The discrepancies between tests and grades are defined as the common-grade dimension, a dimension common to the subject grades in Swedish, English and mathematics. Since the common-grade dimension is defined as the discrepancies between grades and tests common to several subject grades and several teachers, it is interpreted as reflecting aspects which are common to several subjects and several teachers. As indicated by previous research, such a general factor not only reflects verbal ability (to a large extent) but also social-behavioral aspects. Hence, the common-grade dimension is similarly interpreted as reflecting both verbal abilities and, to some extent, social-behavioral aspects of importance for school success. The dimensionality of grades also seems to be stable over cohorts. Moreover, the same pattern of dimensionality can be identified in both criterion-referenced and norm-referenced grades.

These different dimensions in grades also contribute to their predictive validity, meaning that both knowledge and skills, and social-behavioral aspects, predict future school success. Furthermore, the dimensionality contributes to explaining the predictive power of both criterion-referenced and norm-referenced grades. This pattern of predictive validity seems possible to explain, at least in part, in terms of the investment of both cognitive and social-behavioral aspects into the acquisition of knowledge and skills (e.g Cattell, 1987; Gustafsson & Carlstedt, 2006), as well as a direct influence from teachers' grading practices. Hence, both cognitive and social-behavioral aspects encapsulated in grades seem to have importance for subsequent achievement and, in part, explain the predictive validity of grades.

The dimensionality of grades can, in part, be related to teachers' grading practices. It has been indicated elsewhere (e.g. Brookhart, 1991, 1993) that teachers take account of social-behavioral aspects in their grading, which could affect grading directly. This direct influence is believed to be reflected in the common-grade dimension, which also contributes to explaining the

predictive validity of grades. However, it could be argued that when teachers take direct account of such aspects it would result in a distortion of the construct of grades and one could then question whether their use for selection purposes would lead to adverse consequences (Messick, 1989). Still, both cognitive and social-behavioral aspects encapsulated in grades seem to have importance for subsequent achievement and, in part, they explain the predictive validity of grades.

Results also indicate that there were differences with respect to gender and parental education in the common-grade dimension. Girls' higher values in the common-grade dimension indicate that is possible to partially explain girls' higher grades as a consequence of direct and indirect influence of social-behavioral aspects, such as interest in schoolwork (e.g. Klapp Lekholm & Cliffordson, 2009). Additionally, there were effects of parental education in the common-grade dimension which can be partially explained by a compensatory grading mechanism in criterion-referenced grades (e.g. Klapp Lekholm & Cliffordson, 2009). However, in the norm-referenced grades the higher values for students with higher educational background is harder to interpret due to the restricted sample.

Support for Cattell's (1987) Investment theory was also found when investigating the influence of *Gf* and *Gc* on the development of knowledge and skills. The results showed a continuous influence of *Gf* on the measures of *Gc* throughout compulsory school. However, no direct influence from *Gf* on the subject-grades in Grade 9 was found, thus offering partial support for Encapsulation theory (Gustafsson & Carlstedt, 2006). These results indicate that the predictive power of school grades is possible to explain in terms of a continuous investment of *Gf* and an influence of *Gc*, as well as a direct and indirect influence of social-behavioral aspects.

11 Swedish summary

Inledning

I Sverige används betyg som urvalsinstrument till såväl gymnasie- som högskolestudier. Hur urval utformas i olika utbildningssystem beror på om det anses mest fördelaktigt att basera det på instrument som mäter kapacitet för studier, tidigare skolprestationer eller på en kombination av flera typer av instrument. Olika typer av urvalsinstrument kan klassificeras enligt ett kontinuum av formbara och kristalliserade förmågor. Instrument som testar begåvning befinner sig på den formbara änden, instrument som prövar kapacitet för studier, såsom Högskoleprovet, något närmre den kristalliserade änden och mått på tidigare skolprestationer, såsom betyg, längst ut på den kristalliserade änden (Lohman, 2004). Oavsett vilken princip urvalet baseras på så förväntas instrumenten vara både reliabla och valida samt ha förmågan att förutsäga studieframgång.

Instrument som testar begåvning och kapacitet för studier har traditionellt betraktats som mer tillförlitliga mått än instrument som testar skolprestationer, när det gäller att förutsäga framtida studieresultat (Geiser & Santelices, 2008). Detta trots att tidigare prestationer och i synnerhet betyg visat på högre prediktiv förmåga än test som mäter kapacitet för studier (e.g. Bowen, Chingos & McPherson, 2009; Cliffordson, 2008; Geiser & Santelices, 2007; Gustafsson, 2003; Gustafsson & Carlstedt, 2006). Däremot finns det fortfarande kunskapsluckor när det gäller att förklara varför betyg har bättre prognosförmåga.

Kognitiva aspekter relaterade till kunskapsutveckling

Distinktionen mellan begåvning och skolprestation ligger nära distinktionen mellan formbar intelligens (*Gf*) och kristalliserad intelligens (*Gc*). *Gf* representerar förmågan till problemlösning/slutledning och abstraktion, medan *Gc* representerar sådana kunskaper som är kulturellt betingade och som man tillägnar sig genom uppväxten (Cattell, 1963). Denna distinktion är välbelagd genom forskning och den hierarkiska *Gf-Gc* modellen (Horn & Cattell, 1966), som har faktorer av olika generalitet på två nivåer, anses vara en

bra modell över mänsklig intelligens. På den första nivån finns ett antal faktorer som är smalare till sin natur och på den andra nivån finns ett mindre antal breda faktorer, av vilka de viktigaste är *Gf* och *Gc*. Dock finns ingen generell faktor (*g*) på den tredje nivån i Horn och Cattells (1966) modell. Den generella faktor som tidigare identifierats av Spearman bröts istället ner till *Gf* och *Gc*. Carroll (1993) fann, i sin omfattande empiriska studie, ett starkt stöd för Cattell och Horns modell, men identifierade dessutom en generell faktor på den tredje nivån. Denna integrerade modell benämns Cattell-Horn-Carroll modellen (CHC-modellen). Även om CHC-modellen har starkt empiriskt stöd så finns det en konflikt i huruvida en generell faktor skall definieras på den tredje nivån. En lösning på detta problem har dock beskrivits av Gustafsson (1984) som menar att Spearmans *g* och Cattells *Gf* definitionsmässigt är väldigt lika. I en serie empiriska analyser fann också Valentin-Kvist och Gustafsson (2008) att *g* och *Gf* har en perfekt relation i homogena urval och att de därför kan definieras som en och samma faktor.

Denna perfekta relation mellan *g* och *Gf* ger också indirekt stöd för Investmentteorin som utvecklats av Cattell (1971, 1987). Investmentteorin beskriver utvecklingen av intellektuella förmågor genom att beskriva den kausala relationen mellan *Gf* och *Gc*. Investmentteorin föreslår att individuella differenser i *Gc* beror på skillnader i *Gf* och att *Gf* är en generell förmåga som är av betydelse vid inlärning av komplexa områden såsom läsning och matematik. Cattell (1987) framhöll dock att även faktorer som motivation, intresse och undervisningens kvalitet är viktigt för kunskapstillägnandet.

I Gustafssons och Carlstedts (2006) utvidgning av Investmentteorin läggs större vikt vid *Gc* som en prediktor för kunskaper och färdigheter. I deras Inkapslingsteori förklaras den prediktiva förmågan hos *Gc* som att all information från *Gf* är inkapslad i mått på *Gc* som också innehåller information om individuella differenser avseende sådana kunskaper och färdigheter som är av betydelse för fortsatt lärande. De menar också att all information från *Gc* i sin tur är inkapslad i informationen från betyg, men att betygen därutöver omfattar ytterligare aspekter av kunskaper och färdigheter som är viktiga för studieframgång. Dessutom är betyg i större utsträckning influerade av aspekter såsom motivation, intresse och studievanor (Gustafsson & Carlstedt, 2006).

Även Ackerman (1996) framhäver vikten av *Gc* som prediktor i PPIK-teorin (Process, Personlighet, Intressen, och Kunskap). Han lyfter dock i än starkare grad fram betydelsen av personlighet och intressen i tillägnandet och

underhållet av kunskap. PPIK-teorin har utvecklats som en förklaringsmodell till vuxnas intellektuella utveckling och ligger mycket nära Investment-teorin men med en bredare definition av *Gc* och ett tydligare fokus på personlighetsfaktorer.

Validitetsaspekter relaterade till betygsättning

Enligt Messicks (1989) validitetsteori är begreppsvaliditeten den överordnade validitetsaspekten. Vidare menar han att en förutsättning för validitet är att både de teoretiska perspektiven och de empiriska resultaten är i linje med de tolkningar som görs av ett mått som exempelvis betyg. När ett mått används exempelvis i urvalssyfte skall också konsekvenserna av detta vara i linje med de tolkningar som gjorts. Det finns två olika typer av hot mot validiteten; begreppsirrelevant varians och underrepresentation av begreppet. Irrelevant varians innebär att ett mått är för brett och mäter aspekter som inte är en del av det begrepp man vill mäta. Underrepresentation innebär att måttet är för smalt och att bara några aspekter som ingår i begreppet mäts (Messick, 1989).

Tidigare forskning indikerar att det kan finnas problem med både irrelevant varians och underrepresentation hos betyg. Betyg har visat sig vara flerdimensionella och att de förutom kunskaper och färdigheter också mäter sociala och beteendemässiga aspekter som inte kan relateras till elevernas kunskaper (Brookhart, 1993; Cizek et al. 1995; Klapp Lekholm & Cliffordson, 2008; 2009; Randall & Englehard, 2010). Flera sociala och beteendemässiga aspekter har visat sig vara viktiga för lärare när de överväger elevers betyg, såsom ansträngning (Brookhart, 1993) och engagemang i klassrummet (McMillan, 2002). Brookhart (1993) menar att denna typ av betygsättning görs då lärare är medvetna om betygens konsekvenser och att det därför kan anses som mer rättvist eller viktigt att väga in även sådana aspekter.

De nationella proven eller andra typer av standardiserade test brukar användas när exempelvis betygens kvalitet ska bedömas. Resultaten i sådana studier visar att överensstämmelsen mellan prov och betyg är stor men att det också finns tydliga diskrepanser och att variationen mellan lärare är stor (t.ex. Skolverket, 2009). Att endast beakta resultat från standardiserade test i betygsättningen kan dock leda till en underrepresentation av betygen, då proven inte mäter alla delar i kursplanen.

Prediktiv validitet

Även om begreppsvaliditeten är den övergripande formen av validitet är också den prediktiva validiteten viktig i urvalssammanhang. Mått på tidigare skolprestationer såsom kunskapsprov och betyg har visat sig ha högre prognosförmåga än mått på *Gc*, som i sin tur har bättre prognosförmåga än mått på *Gf* (Gustafsson & Carlstedt, 2006). En delförklaring till detta kan vara att betyg återspeglar både *Gf* och *Gc* och ett bredare spektra av kunskaper och färdigheter. Det är också ett rimligt antagande att sociala och beteendemässiga aspekter såsom intresse, motivation, studievanor och självdisciplin kan vara av betydelse för studieframgång och därmed vara en möjlig förklaring till betygens prognosförmåga.

Forskning har visat att den goda prognosförmågan gäller både normrelaterade betyg och kriterierelaterade betyg, vilka båda har bättre prognosförmåga än Högskoleprovet som mäter kapacitet för studier. Detta innebär att oavhängigt betygens syften och konstruktion så har de högre prognosförmåga än mått på kapacitet för studier (Cliffordson, 2004b; 2008). Betyg har också visat sig kunna förutsäga en rad andra resultat såsom jobbprestationer (Roth, BeVier, Switzer & Schippman, 1996), inkomster på lång sikt och produktivitet (Miller, 1998). Dåliga betyg och skolmisslyckanden har också funnits vara relaterat till dålig mental hälsa (Gustafsson et al., 2010).

Syfte

Syftet med avhandlingen är att bättre förstå betygens prediktiva validitet genom att undersöka inflytandet av kognitiva och sociala och beteendemässiga aspekter på betyg och betydelsen av dessa aspekter för betygens prognosförmåga. Ytterligare ett syfte är att undersöka stabiliteten i dimensionaliteten hos betyg samt möjliga skillnader med avseende på kön och utbildningsbakgrund. Investment-teorin, Inkapslingsteorin och PPIK-teorin används för att tolka inflytandet av kognitiva faktorers betydelse för betygens prediktiva validitet. Tidigare forskning kring betygens dimensionalitet och lärares betygsättning används i tolkningen av inflytandet av sociala och beteendemässiga aspekter. Vidare används Messicks validitetsmodell för att diskutera dimensionaliteten i betygen och konsekvenserna av att använda betyg vid urval. Avhandlingen utgörs av tre empiriska studier och en kappa.

Metod

Data

Data för de tre studierna hämtades från Gothenburg Longitudinal Database (GOLD) och Utvärdering Genom Uppföljning (UGU). GOLD innehåller registerdata för alla individer i Sverige födda mellan 1972 och 1992. UGU är ett longitudinellt projekt som bygger på 10-procentiga riksrepresentativa urval för tio födelsekohorter. Studie I baseras på populationsdata för tre kohorter födda 1987, 1988 och 1989 och information om betyg från årskurs 9 och gymnasieskolan, nationella prov, kön och utbildningsbakgrund. Studie II baseras på ett selekterat urval om ca 4000 individer födda 1972. Då betygen i gymnasieskolan inte är jämförbara mellan olika linjer ingick endast de elever som gått en studieförberedande linje. Information om betyg från årskurs 9 och gymnasieskolan, standardprov, kön och utbildningsbakgrund ingick i analysen. Studie III baseras på ett 10-procentigt riksrepresentativt urval om ca 9000 individer. Information om betyg från årskurs 9, standardprov, kunskapsprov i matematik och kognitiva test användes.

Analysmetod

För samtliga analyser användes strukturell ekvationsmodellering (SEM). Med hjälp av SEM kan man skapa latenta variabler som bygger på samvariation mellan ett flertal indikatorer. Latenta variabler kan representera teoretiska begrepp som inte är direkt observerbara. Genom SEM är det också möjligt att definiera relationer mellan latenta variabler vilka dessutom kan vara både oberoende och beroende. Trovärdigheten hos de modeller som skapas diskuteras i relation till teori och tidigare forskning, men testas också gentemot data där olika statistiska mått används för att utvärdera modellerna. Alla analyser i studien är gjorda med hjälp av Mplus (Muthén & Muthén, 2007-2012) inom STREAMS (Gustafsson & Stahl, 2005).

Ett problem när man analyserar data, i synnerhet inom utbildningsområdet, är att data ofta har en hierarkisk struktur (Gustafsson, 2009). Elever är klustrade i klasser och i skolor och individer som tillhör samma kluster tenderar vara mer lika varandra än elever från andra kluster. För att ta hänsyn till detta kan man modellera på flera nivåer, eller, som i föreliggande studier, genom att använda ”complex”-funktionen i Mplus som hanterar sådana klustereffekter.

Resultat

Studie I

Huvudsyftet var att undersöka stabiliteten i dimensionaliteten i kriterierelaterade betyg och de olika dimensionernas betydelse för prognosförmågan.

Resultaten visade att kriterierelaterade betyg var flerdimensionella och att denna dimensionalitet var stabil över kohorter. Betygen mätte både kunskapsrelaterade dimensioner och en dimension som var gemensam för alla betyg och alla lärare, vilken av Klapp Lekholm och Cliffordson (2009) kunde relateras till elevers intresse och föräldrarnas engagemang i sina barns studier. Både de kunskapsrelaterade och den gemensamma dimensionen bidrog till att förutsäga studieframgång i gymnasieskolan, vilket ger stöd åt att flerdimensionaliteten delvis kan förklara betygens prognosförmåga.

Resultaten visade också att det fanns små inkrementella effekter med avseende på kön på gymnasiebetygen, utöver de skillnader som redan medierats genom grundskolebetygen. Det fanns också skillnader i den gemensamma dimensionen till flickors fördel och till fördel för elever med lägre utbildningsbakgrund.

Studie II

Huvudsyftet var att på motsvarande sätt undersöka dimensionaliteten och prognosförmågan för de normrelaterade betygen.

Resultaten visar att även normrelaterade betyg är flerdimensionella och att de uppvisar samma struktur som de kriterierelaterade betygen. Den gemensamma dimensionen var starkare och bidrog mer till de normrelaterade betygens prognosförmåga än vad som identifierades för de kriterierelaterade betygen. Detta resultat är något av en paradox då standardproven ansågs vara mer styrande för betygssättningen än de nationella proven. En möjlig förklaring kan finnas i att standardprovens framförallt var styrande på klassnivå och att läraren därmed hade stor frihet att rangordna individer inom klassen.

Det fanns små inkrementella skillnader på gymnasiebetygen utöver de skillnader som redan medierats genom grundskolebetygen, både vad gäller kön och utbildningsbakgrund. I likhet med resultaten för de kriterierelaterade betygen, fanns det skillnader i den gemensamma dimensionen till flickors fördel. Däremot var skillnaderna i den gemensamma dimensionen till fördel för elever med högre utbildningsbakgrund, vilket skiljer sig från resultaten för

de kriterierelaterade betygen. Denna fördel för elever med högre utbildningsbakgrund kan med stor sannolikhet förklaras av att urvalet var selekterat och att dessa elever oavsett bakgrund har den press eller det stöd hemifrån som krävs för att lyckas i sina studier.

Studie III

I Studie III testades framförallt Cattells (1971, 1987) Investmentteori, d.v.s. inflytandet av *Gf* och *Gc* på utvecklingen av kunskaper och färdigheter mätta genom betyg i årskurs 9.

Resultatet visar att det finns ett fortgående inflytande från *Gf* på utvecklingen av kunskaper och färdigheter, om än betydligt starkare i de tidiga skolåren. Detta kan tolkas som att barn i sin skolgång fortlöpande möter uppgifter av ökande komplexitet, vilka kräver ytterligare investering av *Gf*. Inga direkta effekter av *Gf* fanns dock på betygen i årskurs 9, vilket ger visst stöd åt Gustafssons och Carlstedts (2006) Inkapslingsteori. Det fanns dock substantiella indirekta effekter från *Gf* genom *Gc* på betygen, vilket är i linje med Cattells (1971, 1987) Investmentteori som hävdar att *Gf* är den drivande faktorn bakom tillägnandet av kunskaper och färdigheter.

Diskussion och slutsatser

Resultatet visar att både kriterie- och normrelaterade betyg är flerdimensionella och att flerdimensionaliteten är stabil över kohorter. Betyg återspeglar både ämnesspecifika dimensioner och en dimension som är gemensam för olika ämnen och lärare. Klapp Lekholm och Cliffordson (2008, 2009) identifierade en motsvarande gemensam dimension som var relaterad till intresse och föräldraengagemang. Tidigare forskning har också indikerat att lärare tar hänsyn till sociala och beteendemässiga aspekter när de sätter betyg (Annerstedt & Larsson, 2010; Brookhart, 1991; Randall & Englehard, 2010), exempelvis av kompensatoriska orsaker (Klapp Lekholm & Cliffordson, 2008). Identifikationen av den gemensamma dimensionen tolkas här delvis som en representation av sådana sociala och beteendemässiga aspekter som har ett direkt inflytande på betyg. Vid en genomgång av kursplanerna inom Lpo94 fanns det också gemensamma aspekter som var av verbal art, varför den gemensamma dimensionen också kan tolkas som en reflektion av verbala aspekter. Denna samstämmighet kan inte på samma sätt identifieras i Lgr80.

Ett sådant direkt inflytande från sociala och beteendemässiga aspekter på betyg innebär att betygen till viss del återspeglar irrelevant varians (Messick, 1989), vilket i sin tur kan förvanska tolkningen av betygen. Det är dock viktigt att påpeka att läraren står inför en svår uppgift att balansera de fostrande målen i läroplanen, verkligheten i klassrummet och uppgiften att motivera eleverna med att sätta betyg. Om det finns implicita krav som ingår i betygssättningen får det konsekvenser för tolkningarna av betygen och för individer när de används som urvalsinstrument.

Betygens prognosförmåga

Både de ämnesspecifika dimensionerna och den gemensamma dimensionen bidrar till att förklara betygens prognosförmåga. Den gemensamma dimensionen indikerar att ett direkt inflytande av sociala och beteendemässiga aspekter bidrar till att förklara betygens prognosförmåga. Det finns sannolikt också ett indirekt inflytande av motsvarande aspekter som bidrar till prognosförmågan, vilka kan relateras till den investering av personlighetsrelaterade faktorer som görs i lärandet (t.ex. Gustafsson & Carlstedt, 2006).

Resultaten visar också att investeringar av kognitiva aspekter, d.v.s. av både *Gf* och *Gc*, bidrar till att förklara betygens prognosförmåga. Resultaten stödjer de antaganden som görs i Cattells (1987) Investmentteori, d.v.s. att *Gf* har ett fortgående inflytande på utvecklingen av kunskaper och färdigheter genom hela grundskolan, både i form av direkta och indirekta effekter. I föreliggande studie var inflytandet från *Gf* störst i årskurs 3, men det fanns även direkta effekter i årskurs 6 och 9, om än betydligt svagare, vilket skulle kunna förklaras med att elever fortlöpande möter uppgifter av ökande komplexitet i de tidiga skolåren som kräver ytterligare investering av *Gf*. Att resultaten inte var helt i linje med Gustafssons och Carlstedts (2006) Inkapslingsteori kan bero på att individerna i föreliggande studie är i grundskoleåldern medan de i Gustafssons och Carlstedts (2006) var unga vuxna.

Sammanfattningsvis visar resultaten att betyg, i linje med Cattells (1987) Investmentteori, återspeglar både *Gf* och *Gc*. Resultaten talar också för att de återspeglar ett vidare spektrum av kunskaper och färdigheter och sociala och beteendemässiga aspekter. Dessa resultat kan rimligen förklara varför betyg har högre prediktiv validitet än mått på begåvning eller kapacitet för studier som exempelvis Högskoleprovet. Betygens prediktiva validitet kan således

delvis förklaras i termer av investeringar av kognitiva förmågor (Cattell, 1987) och personlighetsaspekter (Ackerman, 1996; Cattell, 1987), samt ett direkt inflytande av verbala och sociala och beteendemässiga aspekter på betygen. Dessutom mäter betyg prestationer över en lång period och de baseras på en mängd olika sorters bedömningar som är gjorda av olika lärare i olika ämnen, vilket kan vara en ytterligare bidragande förklaring till den prediktiva validiteten.

Betydelse av kön och föräldrautbildning

Flickor befanns ha en högre nivå på den gemensamma dimensionen, vilket är samstämmigt med Klapp Lekholm och Cliffordsons (2009) resultat, där flickors högre nivå på den gemensamma faktorn till stor del medierades av deras generella intresse för skolarbete. I deras studie fanns det också skillnader i de ämnesspecifika dimensionerna som till stor del medierades av intresse för respektive ämnet. Således förefaller flickors högre betyg delvis vara ett resultat av både en indirekt och en direkt effekt av intresse och föräldraengagemang.

Det fanns också effekter av utbildningsbakgrund i den gemensamma dimensionen. I det kriterierelaterade betygssystemet hade elever med lägre utbildningsbakgrund högre nivå på den gemensamma dimensionen, medan elever med högre utbildningsbakgrund hade högre nivå på den gemensamma dimensionen i det normrelaterade betygssystemet. Dessa resultat kan ha flera förklaringar, varav en är att den gemensamma dimensionen som identifierats för kriterierelaterade betyg delvis är ett uttryck för en kompensatorisk betygssättning, d.v.s. att elever med lägre resultat kompenseras genom att få extra betalt för att de anstränger sig (Klapp Lekholm & Cliffordson, 2009). Det är möjligt att denna kompensatoriska effekt är större inom det kriterierelaterade systemet då där finns en godkäntgräns och att sätta ett icke-godkänt betyg kan få större konsekvenser för eleven än att få ett lågt betyg i det normrelaterade systemet. Resultatet för de normrelaterade betygen kan förklaras av begränsningen av urvalet till elever som läste studieförberedande program, vilka med stor sannolikhet, oavsett bakgrund, har det stöd hemifrån som krävs för att de skall lyckas med sina studier.

Begränsningar och fortsatt forskning

Studiernas (I och II) design bygger på antagandet att de nationella proven/standardproven enbart återspeglar ämnesspecifika kunskaper. Det är

dock möjligt att de, liksom betygen, även återspeglar andra aspekter. Om så är fallet är den gemensamma dimensionen i föreliggande studier underskattad.

Det är troligt att även andra ämnen skulle uppvisa motsvarande mönster och därmed ytterligare bidra till styrkan i den gemensamma dimensionen. Detta var dock inte möjligt då nationella prov i ytterligare ämnen inte fanns tillgängliga vid tiden för studierna. I och med att de nationella proven nyligen har utökats till att omfatta fler ämnen ges dock framtida möjligheter att undersöka detta. Det skulle också varit intressant att undersöka betydelsen av personlighetsfaktorer på utvecklingen av kunskaper och färdigheter. Betydelsen av sådana faktorer framhålls av Ackerman (1996) som viktiga i vuxnas kunskapstillägnande men även Cattell (1987) påtalar deras betydelse.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence, 22*(2), 227-257.

Ackerman, P. L. (2000). Domain-specific Knowledge as the "Dark Matter" of Adult Intelligence: *Gf/Gc*, Personality and Interest Correlates. *Journal of Gerontology: Psychological Sciences, 55*(2), 69-84. doi: 10.1093/geronb/55.2.P69

Ackerman, P. L., & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology, 98*(2), 366-381. doi: 10.1037/0022-0663.98.2.366.

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of Individual Differences and Gender Differences in Knowledge. *Journal of Educational Psychology, 93* (4), 797-825.

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

Allen, J. D. (2005). *Grades as valid measures of academic achievement of classroom learning*. The Clearing House, 78(5), 218 – 223.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545-557.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). *Personality, Psychology and Economics.*

Andersson, A. (1998). The dimensionality of the leaving certificate in the Swedish compulsory school. *Scandinavian Journal of Educational Research, 42*(1), 25-40.

Andersson, H. (1991). *De relativa betygens uppgång och fall. En beskrivning och tematisk analys av framväxten och avvecklingen av det relativa betygssystemet.* Educational Reports, Nr 32. Department of Education, Umeå University.

Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are…': Grading in Swedish PEH – problems of validity, comparability and fairness. *European Physical Education Review, 16*(2), 97-115.

Atkinson, R. C. (2001). Achivement versus Aptitude in College Admissions. *Issues in Science and Technology Online*. Winter, 2001.

Atkinson, R. C. (2004). Achievement versus aptitude in college admission. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 15–23). New York: Routledge Falmer.

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. Reserch & Occasional Paper Series: CSHE.4.09. Center for Studies in Higher Education. University of California, Berkley

Beier, M. E., & Ackerman, P. L. (2001). Current-events knowledge in adults: An investigation of age, intelligence, and nonability determinants. *Psychology and Aging, 16*(4), 615-628.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising Results in a Real-World Learning Environment. *Psychology and Aging, 20* (2), 341-355.

Björnsson, M. (2005). Kön och skolframgång, tolkningar och perspektiv. [Gender and school success, interpretations and perspectives]. Authority for School development. Report no. 13. Stockholm: Liber distribution.

Bloom, B. S. (1976). *Human characteristics and School learning*. New York: McGraw-Hill Book Company.

Bowens, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton: Princeton University Press.

Bowers, A. J. (2008). Reconsidering grades as data for decision making: more than just academic knowledge. *Journal of Educational Administration, 47*(5), 609-629.

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159.

Brennan, L. R. (2006) (ed). *Educational measurement 4th ed.* Westport, CT: Praeger.

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. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy and Practice, 20*(1), 69-90.

Brown, T, A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Böhlmark, A., & Holmlund, H. (2011). *20 år med förändringar i skolan: Vad har hänt med likvärdigheten?* [20 years of changes in school: What has happened to the equality?]. Stockholm: SNS Förlag.

Camara, W. J., & Echternacht, G. (2000). The SAT I and High School Grades: Utility in Predicting Success in College. Research Notes. The College Board. Office of Research and Development.

Carlstedt, B., & Gustafsson, J. E. (2005) Construct validation of the Swedish Scholastic Aptitude Test by means of the Swedish Enlistment Battery. *Scandinavian Journal of Psychology, 46*(1), 31-42. Doi: 10.1111/j.1467-9450.2005.00432x

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Camebridge: Camebridge University Press.

Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (red): *Handbook of human intelligence*. New York: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22.

Cattell, R.B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. New York: North-Holland.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995). Teachers' Assessment Practices: Preparation, Isolation, and the Kitchen Sink. *Educational Assessment, 3*(2), 159-179.

Cliffordson, C. (2004a). Betygsinflation i de målrelaterade gymnasiebetygen [Inflation in the criterion-referenced grades from upper secondary school]. *Pedagogisk Forskning i Sverige, 9*(1), 1-14.

Cliffordson, C. (2004b). De målrelaterade gymnasiebetygens prognosförmåga [The predictive validity of the criterion-referenced grades from upper secondary school]. *Pedagogisk Forskning i Sverige, 9*(2), 129-140.

Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment, 13*(1), 56-75.

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

Cronbach, L. J. (1971). Essentials of psychological testing (3th ed.). New York: Harper & Row

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52* (4), 281-302.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge Grading: Endorsed by Students and Teachers Alike. *Applied Measurement in Education, 12*(1), 53-72.

Cureton, E. E. (1951). Validity. In: Lindquist, E. F. (Ed.), *Educational measurement* (1st ed., pp. 621-694). Washington, DC: American Council on Education.

Dennis, I. (2007). Halo Effects in Grading Student Projects. *Journal of Applied Psychology, 92* (4), 1169-1176.

de Jong, P.F., & van der Leij, A. (2003). Developmental changes in the Manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95*, 22-40.

Ds 1990:60. *Betygens effekter på undervisningen*. [The effects of grades on the teaching] Stockholm: Ministry of Education.

Eklöf, H., & Nyroos, M. (2013). Pupil perceptions of national tests in science: perceived importance, invested effort, and test anxiety. *European Journal of Psychology of Education, 28* (2), 497-510.

Emanuelsson, I., & Fischbein, S. (1986) Vive la difference? A Study on Sex and Schooling. *Scandinavian Journal of Educational Research, 30*, 71-84.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2001). The Relations Between Measures of Cattell-Horn-Carroll (CHC) Cognitive Abilities and Reading Achievement During Childhood and Adolescence. *School Psychology Review, 31* (2), 246-262.

Fleming, J. (2002). Who Will Succeed in College? When the SAT Predicts Black Students' Performance. *The Review of Higher Education, 25* (3), 281-296.

Fleming, J., & Garcia, N. (1998). Are standardized tests fair to African Americans? Predictive validity of the SAT in Black and White institutions. *The Journal of Higher Education, 69* (5), 471-495.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in The Schools, 40* (2), 155-171.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The Relative Contributions of the Cattell-Horn-Carroll Cognitive Abilities in Explaining

Writing Achievement During Childhood and Adolescence. *Psychology in The Schools, 45* (2), 132-144.

Francis, D. J., Fletcher, J. M., Maxwell, S. E. & Satz, P. (1989). A structural model for developmental changes in the determinants of reading achievement. *Journal of Clinical Child Psychology, 18*, 44-51.

Geiser, S., & Santelices, V. (2007). Validity Of High-School Grades In Predicting Student Success Beyond The Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. Center for Studies in Higher Education (June 2007).

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*(8), 519-521.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179-203.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence Vol. 4* (pp. 35–71). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Gustafsson, J.-E. (2003). The predictive validity of grades and aptitude tests in higher education. Paper presented at the annual conference of the Association for Education Assessment - Europe, Lyon, 6 – 8 November, 2003, France.

Gustafsson, J.-E. (2006). *Barns utbildningssituation – bidrag till ett kommunalt barnindex*. [Childrens' educational situation – contributions to a municipal childrens' index] Stockholm: Save the children.

Gustafsson, J. -E. (2009). Strukturella ekvationsmodeller. [Structural equation models]. In: Djurfeldt, G. & Barmark, M. (2009). Statistisk verktygslåda – multivariat analys. [Statistical tool box –Multivariate analysis]. Lund: Studentlitteratur.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Educational Assessment, 13*, 56-75.

Gustafsson, J.-E., & Carlstedt, B. (2006, August). Abilities and grades as predictors of achievement: The encapsulation theory. Paper presented at the symposium "The investment theory of intelligence: New evidence, new challenges" at the annual meeting of the American Psychological Associations, New Orleans.

Gustafsson, J.-E., & Yang-Hansen, K. (2009), Resultatförändringar i svensk grundskola.[Changes of results in Swedish compulsory school] In: National

Agency for Education (2009) *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer*. [What affects the results in Swedish compulsory school? An overview on the importance of different factors]. Stockholm: National Agency for Education.

Gustafsson, J. –E., Allodi, M. W., Alin Åkerman, B., Eriksson, C., Eriksson, L., Fischbein, S., Granlund, M., Gustafsson, P., Ljungdahl, S., Ogden, T., & Persson, R.S. (2010). *School, Learning and Mental Health: a systematic review*. Stockholm: The Royal Swedish Academy of Sciences, The Health Committee.

Gustafsson, J. –E., & Erickson, G. (2013). To trust or not to trust? –teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability, 25* (1), 69-87.

Gustafsson, J. –E., Cliffordson, C., & Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter*. [Equal assessment in and of the Swedish school – problems and possibilities]. Stockholm: SNS Förlag.

Harlen, W. (2005). Trusting teachers' judgment: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20*, 245-270.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.

Hinnerich, B. T., & Vlachos, J. (2013). Systematiska skillnader mellan interna och externa bedömningar av nationella prov – en uppföljningsrapport. [Systematic differences between external assessments of national tests – a follow-up report]. Supplement to The School Inspectorate (2013). *Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan*. [The differences are too large. Reassessments of national tests in compulsory school and upper secondary school].

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hoffman, J. L., & Luwitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education, 28* (4), 455-474.

Hoge, R. D., & Coladarci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research, 59* (3), 297-313.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*(5), 253-270.

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In: Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment third edition*. New York: The Guilford Press.

Husén, T. (1986). Genombrottsåren för svensk utbildningsforskning: krafter, idéer, aktörer. [The years of breakthrough for Swedish educational research: forces, ideas, participants] In: Gustafsson, J. E., & Marton, F. (red). *Pedagogikens gränser och möjligheter.* [The limits and possibilities of the science of education]. Lund: Studentlitteratur. (24-49).

Johansson, S., Myrberg, E., & Rosén, M. (2012). Teachers and tests: assessing pupils' reading achievement in primary schools. *Educational Research and Evaluation, 18*(8), 693-711.

Jöreskog, K. G. (1993). Testing Structural Equation Models. In: Bollen, K. A., & Long, S. J. (1993). *Testing Structural Equation Models*. London: Sage Publications.

Kan, K. J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the CHC factor *Gc. Intelligence, 39* (5), 292-302.

Klapp-Lekholm, A. (2008). *Grades and grade assignment: effects of student and school characteristics.* [Doctoral Thesis]. Gothenburg: Acta Universitatis Gothoburgensis.

Klapp-Lekholm, A. (2010). Vad mäter betygen? [What do grades measure?] In Lundahl, C., & Folke-Fichtelius, M. (ed). (2010). *Bedömning i och av skolan.* [Assessment in and of school] p. 129-138. Lund: Studentlitteratur AB

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.

Komarraju, M., Ramsey, A., & Rinella, V., (2012). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline. *Learning and Individual Differences, 24*, 103-109.

Kroksmark, T. (2000). En tankes fall i praktiken – då den målrationella styrningen möter skolan. [A fall of a thought in practice – when the goal oriented steering meets school] In: *Att bedöma eller döma. Tio artiklar om*

bedömning och betygssättning, [To assess or judge. Ten articles on assessment and grading]. (pp. 57-56). Stockholm: National Agency for Education.

Levin, H. M. (2011). The utility and need for incorporating non-cognitive skills into large scale educational assessments. Paper presented at ETS Invitational Conference on International Large-Scale Assessment and to be published by Educational Testing Service.

Lgr11. *2011 års Läroplan för grundskolan, förskoleklassen och fritidshemmet.* [Curriculum for compulsory school, pre-school and after school center] Stockholm: National Agency for Education.

Lgr62. *1962 års Läroplan för grundskolan.* [The curriculum for compulsory school, year 1962]. Stockholm: SO Förlaget

Lgr80. *1980 års Läroplan för Grundskolan.*[Curriculum for compulsory school, year 1980]. Södertälje: Skolöverstyrelsen och Liber UtbildningsFörlaget.

Lgy70. *1970 års Läroplan för gymnasieskolan.* [Curriculum for upper secondary school, year 1970]. Stockholm: Skolöverstyrelsen.

Ljung, B. O. (2000). Standardproven – 53 år i skolans tjänst. [Standard-tests – 53 years in the schools' service]. Report from the PRIM-group nr 17. Stockholm: Elanders Gotab.

Lohman, D. (2004). Aptitude for college: The importance of reasoning tests for minority admissions. In R. Zwick (Ed.). *Rethinking the SAT. The future of standardized testing in university admissions*, (pp. 41-56). New York and London: RoutledgeFalmer.

Lpf94. *1994 års Läroplan för de frivilliga skolformerna.* [Curriculum for upper secondary school, year 1994]. Stockholm: Skolverket.

Lpo94. *1994 års Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet.* [Curriculum for compulsory school, pre-school and after school center]. Stocholm: National Agency for Education.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1–10.

McGrew, K. S., & Knopik, S.N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement. *School Psychology Review, 22* (4), 687-695.

McMillan, J. H. (2001). Essential assessment concepts for teachers and administrators. Thousand Oaks, CA: Corwin Publishing Company.

McMillan, J. H. (2003). Understanding and Improving Teachers' Classroom Assessment Decision Making: Implications for Theory and Practice. *Educational Measurement: Issues and Practice*, Winter 2003, 34-43.

McMillan, J. H., Myran S., & Workman D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*, 203-213.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Third edition, pp 13-103). New York: American Council on Education/Macmillan.

Messick, S. (1990). *Validity of test interpretation and use.* Research report. Princeton: Educational testing service.

Miller, S. R. (1998). High school grades as a signal of human capital. *Educational Evaluation and Policy Analysis, 20* (4), 299-311.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30* (109).

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, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.

Mårell-Olsson, E. (2012). *Att göra lärandet synligt? Individuella utvecklingsplaner och digital dokumentation.* [Doing learning visible? Individual development plans and digital documentation]. Academic thesis. Department of applied educational science. Umeå University.

National Agency for Education (2004a). *Likvärdig bedömning och betygssättning. Skolverkets Allmänna råd 2004.* [Equal assessment and grading. General advice from National Agency for Education, 2004] Stockholm: National Agency for Education.

National Agency for Education (2004b). *Prövostenar i prakiken. Grundskolans nationella provsystem i ljuset av användares synpunkter.* [Tests in practice. The national testing system in compulsory school in the light of the viewpoints of its users]. Stockholm: National Agency for Education.

National Agency for Education (2006). Könsskillnader i måluppfyllelse och utbildningsval. [Gender differences in goal attainment and educational choice]. Stockholm: National Agency for Education.

National Agency for Education (2007). *Provbetyg – Slutbetyg – Likvärdig bedömning? En statistisk analys av sambandet mellan provbetyg och slutbetyg i*

grundskolan, 1998-2006. [Test grades – Final grades – Equal assessment? A statistical analysis of the relation between test grades and final grades in compulsory school, 1998-2006] Stockholm: National Agency for Education.

National Agency for Education (2009a). *Författningskommentarer för likvärdig bedömning och betygsättning*. [Constitutional commentaries for equal assessment and grading]. Stockholm: National Agency for Education.

National Agency for Education (2009b). *Bedömaröverensstämmelse vid bedömning av nationella prov*. [Inter-rater agreement in assessment of national tests]. Stockholm: National Agency for Education.

National Agency for Education (2012a). *Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolans årskurs 9*. [Accounts of the commission on the divergence between test results and grades in year 9 of compulsory school]. Dnr 75-2012:311. Stockholm: National Agency for Education.

National Agency for Education (2012b). *Nationella prov – Skolverkets synpunkter i anledning av Skolinspektionens omrättning*. [National tests – The viewpoints of the National Agency for Education with respect to the re-assessments of the School Inspectorate]. Dnr. 73-2012:1512. Stockholm: National Agency for Education.

National Agency for Education (2012c). *Likvärdig utbildning i svensk grundskola? En kvantitativ analys av likvärdighet över tid*. [Equal education in Swedish compulsory school? A quantitative analysis of equality over time]. Report 374. Stockholm: National Agency for Education.

National Agency for Education (2013). *Utbildningsresultat på riksnivå*. [Educational results on national level]. Report no 382. Stockholm: National Agency for Education.

Nisbett, R. E., & Wilson, T. DC. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology, 35* (4), 250-256.

Nycander, M. (2006). Pojkars och flickors betyg – en statistisk undersökning. [Boys' and girls' grades – a statistical investigation]. Uppsala University: Institution for Teacher Education.

Nyström, P. (2004). *Rätt mätt på prov. Om validering av bedömningar i skolan*. Dissertation. Department of Education. Umeå University.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6*(1), 1-9.

Postlethwaite, B. E. (2011). *Fluid ability, crystallized ability, and performance across multiple domains: a meta-analysis*. Dissertation, University of Iowa.

Prop. 2008/09:87. Tydligare mål och kunskapskrav – nya läroplaner för skolan. [Clearer goals and knowledge demands – new curricula for school.] Government bill.

Randall, J., & Engelhard, G. (2009a). Differences between Teachers' Grading Practices in Elementary and Middle schools. *The Journal of Educational Research, 102* (3), 175-186.

Randall, J. & Engelhard, G. (2009b). Examining Teacher Grades Using Rasch Measurement Theory. *Journal of Educational Measurement, 46* (1), 1-18.

Randall, J. & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*, 1372-1380.

Reuterberg, S-E., & Svensson, A. (2000). Köns- och socialgruppsskillnader i matematik - orsaker och konsekvenser. [Gender- and social group differences in mathematics – causes and consequences]. IPD-report nr 2000:20. Mölndal: University of Gothenburg, Institution för education and didactics.

Richardsson, G. (2004). *Svensk Utbildningshistoria.* [Swedish Educational History]. Stockholm: Studentlitteratur.

Rosander, P. (2013). *The importance of personality, IQ and learning approaches: Predicting academic performance.* Academic Thesis, Department of Psychology, Lund University.

Roth, P. L., BeVier, C. A., Switzer, F. S., & Schippman, J. S. (1996). Meta-Analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81* (5), 548-556.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7* (2), 147–177. doi:10.1037/1082-989X.7.2.147.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll Model of Intelligence. In: Flanagan, D. P. & Harrison, P. L. (ed). *Contemporary intellectual assessment 3 ed.* New York: The Guilford Press.

School Inspectorate (2012). *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år. Redovisning av regeringsuppdrag.* [Equal for all? Re-assessment of national tests in compulsory school and upper secondary school during three years. Account of government commission]. Dnr. U2009/4877/G

SOU 1942:11. *Betänkande med utredning och förslag angående betygssättningen i folkskolan.* [Comission report on investigation and suggestions concerning grading in compulsory school] Stockholm: Ministry of Education.

SOU 1945:45. *Skolans betygssättning.* [Grading in school] Stockholm: Ministry of Education.

SOU 1963:42. *Ett nytt gymnasium*. [A new upper secondary school]. Stockholm: Ministry of Education.

SOU 1977:9. *Betygen i skolan. Betänkande av 1973 års betygsutredning.* [Grades in school. Report by 1973 years grade commission]. Stockholm: Gotab.

SOU 2007: 28. *Tydliga mål och kunskapskrav i grundskolan. Förslag till nytt mål- och uppföljningssystem.* [Clear goals and knowledge demands in compulsory school. Suggestions on a new goal- and evaluation system]. Stockholm: Statens offentliga utredningar.

Stemler, S. E. 2004. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability [Electronic version]. *Practical Assessment, Research & Evaluation, 9*. Retrieved 11 November 2013 from: http://PAREonline.net/getvn.asp?v=9&n=4.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology,* 104, 743-762.

Svensson, A. (1971). *Relative Achievement. School performance in relation to intelligence, sex and home environment*. Göteborg Studies in Educational Sciences, 6.

Svensson, A. (2001). Består den sociala snedrekryteringen? Elevernas val av gymnasieprogram hösten 1998. [Does the socially skewed recruitment remain? Students' choice of upper secondary school track fall 1998]. *Pedagogisk Forskning i Sverige 6*, (3) 161 - 172.

Svensson, A. (2008). Har dagens tonåringar sämre studieförutsättningar? En studie av förskjutningar i intelligenstestresultat från 1960-talet och framåt. [Do the teenagers of today have poorer prerequisites for studies? A study of displacements in results of intelligence tests from the 1960's until today.] *Pedagogisk Forskning i Sverige, 13* (4), 258–277.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal, 31*(2), 231-262.

The Swedish Education Act (2010). *Svensk författningssamling 2010:800* [Electronic Resource] Retrieved from:

http://www.riksdagen.se/sv/DokumentLagar/Lagar/Svenskforfattningssamling/Skollag-2010800_sfs-2010-800/ 3 December 2013.

Tholin, J. (2006). *Att kunna klara sig i okänd natur. En studie av betyg och betygskriterier – historiska betingelser och implementering av ett nytt system.* [Being able to survive in an unknown environment. A study of grades and grade criteria – historical factors and the implementation of a new system]. Academic thesis. Borås: The University college of Borås, Department of Education.

Thorsen, C. (2014). Dimensions of Norm-Referenced Compulsory School Grades and their Relative Importance for the Prediction of Upper Secondary School Grades. *Scandinavian Journal of Educational Research, 58* (2), 127-146, DOI: 10.1080/00313831.2012.705322.

Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation: An International Journal on Theory and Practice, 18*(2), 153-172.

Valentin Kvist, A, & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36* (5), 422-436.

Wernersson, I. (2010). Könsskillnader i skolprestationer – Idéer om orsaker. [Gender differences in school achievement – Ideas on causes]. Report V from the Delegation on equality in school. Stockholm: Statens offentliga utredningar.

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

Wikström, C., & Wikström, M. (2012). *Urval till högre utbildning – Påverkas betygens prediktionsvärde av ålder?* [Selection to higher education – Is the predictive validity of grades affected by age?] Report 2012:21. Uppsala: Institute for Labour Market and Education Policy.

Willingham, W., & Cole, N. (1997). *Gender and Fair Assessment.* London: Lawrence Erlbaum Associates, Publishers.

Woessmann, L. (2004). How equal are educational opportunities? Family background and student achievement in Europe and US. CESifo Working Paper no 1162. Category 4: Labour markets. March 2004.

Woessmann, L., & Peterson, P. E. (2007). *Schools and the equal opportunity problem.* Cambridge: Mass, MIT Press.

Wolming, S., & Wikström, C. (2010). The concept of validity in theory and practice. *Assessment in Education: Principles, Policy & Practice, 17*(2), 117-132.

Yang, Y. (2003). Measuring socioeconomic status and its effects at individual and collective levels: A cross-country comparison. Academic Thesis. Göteborg studies in educational science, Vol.193. Göteborg Sweden: Acta Universitatis Gothoburgensis.

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

Tidigare utgåvor:

Editors: Kjell Härnqvist and Karl-Gustaf Stukát

1. KARL-GUSTAF STUKÁT *Lekskolans inverkan på barns utveckling.* Stockholm 1966

2. URBAN DAHLLÖF *Skoldifferentiering och undervisningsförlopp*. Stockholm 1967

3. ERIK WALLIN *Spelling. Factorial and experimental studies.* Stockholm 1967

4. BENGT-ERIK ANDERSSON *Studies in adolescent behaviour. Project Yg, Youth in Göteborg.* Stockholm 1969

5. FERENCE MARTON *Structural dynamics of learning.* Stockholm 1970

6. ALLAN SVENSSON *Relative achievement. School performance in relation to intelligence, sex and home environment.* Stockholm 1971

7. GUNNI KÄRRBY *Child rearing and the development of moral structure.* Stockholm 1971

Editors: Urban Dahllöf, Kjell Härnqvist and Karl-Gustaf Stukát

8. ULF P. LUNDGREN *Frame factors and the teaching process. A contribution to curriculum theory and theory on teaching.* Stockholm 1972

9. LENNART LEVIN *Comparative studies in foreign-language teaching.* Stockholm 1972

10. RODNEY ÅSBERG *Primary education and national development*. Stockholm 1973

11. BJÖRN SANDGREN *Kreativ utveckling.* Stockholm 1974

12. CHRISTER BRUSLING *Microteaching - A concept in development*. Stockholm 1974

13. KJELL RUBENSON *Rekrytering till vuxenutbildning. En studie av kortutbildade yngre män.* Göteborg 1975

14. ROGER SÄLJÖ *Qualitative differences in learning as a function of the learner's conception of the task.* Göteborg 1975

15. LARS OWE DAHLGREN *Qualitative differences in learning as a function of content-oriented guidance.* Göteborg 1975

16. MARIE MÅNSSON *Samarbete och samarbetsförmåga. En kritisk granskning.* Lund 1975

17. JAN-ERIC GUSTAFSSON *Verbal and figural aptitudes in relation to instructional methods. Studies in aptitude - treatment interactions.* Göteborg 1976

18. MATS EKHOLM *Social utveckling i skolan. Studier och diskussion*. Göteborg 1976

19. LENNART SVENSSON *Study skill and learning.* Göteborg 1976

20. BJÖRN ANDERSSON *Science teaching and the development of thinking.* Göteborg 1976

21. JAN-ERIK PERNEMAN *Medvetenhet genom utbildning.* Göteborg 1977

Editors: Kjell Härnqvist, Ference Marton and Karl-Gustaf Stukát

22. INGA WERNERSSON *Könsdifferentiering i grundskolan.* Göteborg 1977

23. BERT AGGESTEDT & ULLA TEBELIUS *Barns upplevelser av idrott.* Göteborg 1977

24. ANDERS FRANSSON *Att rädas prov och att vilja veta.* Göteborg 1978

25. ROLAND BJÖRKBERG *Föreställningar om arbete, utveckling och livsrytm.* Göteborg 1978

26. GUNILLA SVINGBY *Läroplaner som styrmedel för svensk obligatorisk skola. Teoretisk analys och ett empiriskt bidrag.* Göteborg 1978

27. INGA ANDERSSON *Tankestilar och hemmiljö.* Göteborg 1979

28. GUNNAR STANGVIK *Self-concept and school segregation.* Göteborg 1979

29. MARGARETA KRISTIANSSON *Matematikkunskaper Lgr 62, Lgr 69*. Göteborg 1979

30. BRITT JOHANSSON *Kunskapsbehov i omvårdnadsarbete och kunskapskrav i vårdutbildning.* Göteborg 1979

31. GÖRAN PATRIKSSON *Socialisation och involvering i idrott.* Göteborg 1979

32. PETER GILL *Moral judgments of violence among Irish and Swedish adolescents.* Göteborg 1979

33. TAGE LJUNGBLAD *Förskola - grundskola i samverkan. Förutsättningar och hinder.* Göteborg 1980

34. BERNER LINDSTRÖM *Forms of representation, content and learning.* Göteborg 1980

35. CLAES-GÖRAN WENESTAM *Qualitative differences in retention.* Göteborg 1980

36. BRITT JOHANSSON Pedagogiska samtal i vårdutbildning. Innehåll och språkbruk. Göteborg 1981

37. LEIF LYBECK *Arkimedes i klassen. En ämnespedagogisk berättelse.* Göteborg 1981

38. BIÖRN HASSELGREN *Ways of apprehending children at play. A study of pre-school student teachers' development.* Göteborg 1981

39. LENNART NILSSON *Yrkesutbildning i nutidshistoriskt perspektiv. Yrkesutbildningens utveckling från skråväsendets upphörande 1846 till 1980-talet samt tankar om framtida inriktning.* Göteborg 1981

40. GUDRUN BALKE-AURELL *Changes in ability as related to educational and occupational experience.* Göteborg 1982

41. ROGER SÄLJÖ *Learning and understanding. A study of differences in constructing meaning from a text.* Göteborg 1982

42. ULLA MARKLUND *Droger och påverkan. Elevanalys som utgångspunkt för drogundervisning.* Göteborg 1983

43. SVEN SETTERLIND *Avslappningsträning i skolan. Forskningsöversikt och empiriska studier.* Göteborg 1983

44. EGIL ANDERSSON & MARIA LAWENIUS *Lärares uppfattning av undervisning.* Göteborg 1983

45. JAN THEMAN *Uppfattningar av politisk makt.* Göteborg 1983

46. INGRID PRAMLING *The child's conception of learning.* Göteborg 1983

47. PER OLOF THÅNG *Vuxenlärarens förhållningssätt till deltagarerfarenheter. En studie inom AMU.* Göteborg 1984

48. INGE JOHANSSON *Fritidspedagog på fritidshem. En yrkesgrupps syn på sitt arbete.* Göteborg 1984

49. GUNILLA SVANBERG *Medansvar i undervisning. Metoder för observation och kvalitativ analys.* Göteborg 1984

50. SVEN-ERIC REUTERBERG *Studiemedel och rekrytering till högskolan.* Göteborg 1984

51. GÖSTA DAHLGREN & LARS-ERIK OLSSON *Läsning i barnperspektiv.* Göteborg 1985

52. CHRISTINA KÄRRQVIST *Kunskapsutveckling genom experimentcentrerade dialoger i ellära.* Göteborg 1985

53. CLAES ALEXANDERSSON *Stabilitet och förändring. En empirisk studie av förhållandet mellan skolkunskap och vardagsvetande.* Göteborg 1985

54. LILLEMOR JERNQVIST *Speech regulation of motor acts as used by cerebral palsied children. Observational and experimental studies of a key feature of conductive education.* Göteborg 1985

55. SOLVEIG HÄGGLUND *Sex-typing and development in an ecological perspective.* Göteborg 1986

56. INGRID CARLGREN *Lokalt utvecklingsarbete.* Göteborg 1986

57. LARSSON, ALEXANDERSSON, HELMSTAD & THÅNG *Arbetsupplevelse och utbildningssyn hos icke facklärda.* Göteborg 1986

58. ELVI WALLDAL *Studerande vid gymnasieskolans vårdlinje. Förväntad yrkesposition, rollpåverkan, självuppfattning.* Göteborg 1986

Editors: Jan-Eric Gustafsson, Ference Marton and Karl-Gustaf Stukát

59. EIE ERICSSON *Foreign language teaching from the point of view of certain student activities.* Göteborg 1986

60. JAN HOLMER *Högre utbildning för lågutbildade i industrin.* Göteborg 1987

61. ANDERS HILL & TULLIE RABE *Psykiskt utvecklingsstörda i kommunal förskola.* Göteborg 1987

62. DAGMAR NEUMAN *The origin of arithmetic skills. A phenomenographic approach.* Göteborg 1987

63. TOMAS KROKSMARK *Fenomenografisk didaktik.* Göteborg 1987

64. ROLF LANDER *Utvärderingsforskning - till vilken nytta?* Göteborg 1987

65. TORGNY OTTOSSON *Map-reading and wayfinding.* Göteborg 1987

66. MAC MURRAY *Utbildningsexpansion, jämlikhet och avlänkning.* Göteborg 1988

67. ALBERTO NAGLE CAJES *Studievalet ur den väljandes perspektiv.* Göteborg 1988

68. GÖRAN LASSBO *Mamma - (Pappa) - barn. En utvecklingsekologisk studie av socialisation i olika familjetyper.* Göteborg 1988

69. LENA RENSTRÖM *Conceptions of matter. A phenomenographic approach.* Göteborg 1988

70. INGRID PRAMLING *Att lära barn lära.* Göteborg 1988

71. LARS FREDHOLM *Praktik som bärare av undervisnings innehåll och form. En förklaringsmodell för uppkomst av undervisningshandlingar inom en totalförsvarsorganisation.* Göteborg 1988

72. OLOF F. LUNDQUIST *Studiestöd för vuxna. Utveckling, utnyttjande, utfall.* Göteborg 1989

73. BO DAHLIN *Religionen, själen och livets mening. En fenomenografisk och existensfilosofisk studie av religionsundervisningens villkor.* Göteborg 1989

74. SUSANNE BJÖRKDAHL ORDELL *Socialarbetare. Bakgrund, utbildning och yrkesliv.* Göteborg 1990

75. EVA BJÖRCK-ÅKESSON *Measuring Sensation Seeking.* Göteborg 1990

76. ULLA-BRITT BLADINI *Från hjälpskolelärare till förändringsagent. Svensk speciallärarutbildning 1921-1981 relaterad till specialundervisningens utveckling och förändringar i speciallärarens yrkesuppgifter.* Göteborg 1990

77. ELISABET ÖHRN *Könsmönster i klassrumsinteraktion. En observations- och intervjustudie av högstadieelevers lärarkontakter.* Göteborg 1991

78. TOMAS KROKSMARK *Pedagogikens vägar till dess första svenska professur.* Göteborg 1991

Editors: Ingemar Emanuelsson, Jan-Eric Gustafsson and Ference Marton

79. ELVI WALLDAL *Problembaserad inlärning. Utvärdering av påbyggnadslinjen Utbildning i öppen hälso- och sjukvård.* Göteborg 1991

80. ULLA AXNER *Visuella perceptionssvårigheter i skolperspektiv. En longitudinell studie.* Göteborg 1991

81. BIRGITTA KULLBERG *Learning to learn to read.* Göteborg 1991

82. CLAES ANNERSTEDT *Idrottslärarna och idrottsämnet. Utveckling, mål, kompetens - ett didaktiskt perspektiv.* Göteborg 1991

83. EWA PILHAMMAR ANDERSSON *Det är vi som är dom. Sjuksköterskestuderandes föreställningar och perspektiv under utbildningstiden.* Göteborg 1991

84. ELSA NORDIN *Kunskaper och uppfattningar om maten och dess funktioner i kroppen. Kombinerad enkät- och intervjustudie i grundskolans årskurser 3, 6 och 9.* Göteborg 1992

85. VALENTIN GONZÁLEZ *On human attitudes. Root metaphors in theoretical conceptions.* Göteborg 1992

86. JAN-ERIK JOHANSSON *Metodikämnet i förskollärarutbildningen. Bidrag till en traditionsbestämning.* Göteborg 1992

87. ANN AHLBERG *Att möta matematiska problem. En belysning av barns lärande.* Göteborg 1992

88. ELLA DANIELSON *Omvårdnad och dess psykosociala inslag. Sjuksköterskestuderandes uppfattningar av centrala termer och reaktioner inför en omvårdnadssituation.* Göteborg 1992

89. SHIRLEY BOOTH *Learning to program. A phenomenographic perspective.* Göteborg 1992

90. EVA BJÖRCK-ÅKESON *Samspel mellan små barn med rörelsehinder och talhandikapp och deras föräldrar - en longitudinell studie.* Göteborg 1992

91. KARIN DAHLBERG *Helhetssyn i vården. En uppgift för sjuksköterskeutbildningen.* 1992

92. RIGMOR ERIKSSON *Teaching Language Learning. In-service training for communicative teaching and self directed learning in English as a foreign language.* 1993

93. KJELL HÄRENSTAM *Skolboks-islam. Analys av bilden av islam i läroböcker i religionskunskap.* Göteborg 1993.

94. INGRID PRAMLING *Kunnandets grunder. Prövning av en fenomenografisk ansats till att utveckla barns sätt att uppfatta sin omvärld.* Göteborg 1994.

95. MARIANNE HANSSON SCHERMAN *Att vägra vara sjuk. En longitudinell studie av förhållningssätt till astma/allergi.* Göteborg 1994

96. MIKAEL ALEXANDERSSON *Metod och medvetande.* Göteborg 1994

97. GUN UNENGE *Pappor i föräldrakooperativa daghem. En deskriptiv studie av pappors medverkan.* Göteborg 1994

98. BJÖRN SJÖSTRÖM *Assessing acute postoperative pain. Assessment strategies and quality in relation to clinical experience and professional role.* Göteborg 1995

99. MAJ ARVIDSSON *Lärares orsaks- och åtgärdstankar om elever med svårigheter.* Göteborg 1995

100. DENNIS BEACH *Making sense of the problems of change: An ethnographic study of a teacher education reform.* Göteborg 1995.

101. WOLMAR CHRISTENSSON *Subjektiv bedömning - som besluts och handlingsunderlag.* Göteborg 1995

102. SONJA KIHLSTRÖM *Att vara förskollärare. Om yrkets pedagogiska innebörder.* Göteborg 1995

103. MARITA LINDAHL *Inlärning och erfarande. Ettåringars möte med förskolans värld.* Göteborg. 1996

104. GÖRAN FOLKESTAD *Computer Based Creative Music Making - Young Peoples´ Music in the Digital Age.* Göteborg 1996

105. EVA EKEBLAD *Children • Learning • Numbers. A phenomenographic excursion into first-grade children's arithmetic.* Göteborg 1996

106. HELGE STRÖMDAHL *On mole and amount of substance. A study of the dynamics of concept formation and concept attainment.* Göteborg 1996

107. MARGARETA HAMMARSTRÖM *Varför inte högskola? En longitudinell studie av olika faktorers betydelse för studiebegåvade ungdomars utbildningskarriär.* Göteborg 1996

108. BJÖRN MÅRDÉN *Rektorers tänkande. En kritisk betraktelse av skolledarskap.* Göteborg 1996

109. GLORIA DALL'ALBA & BIÖRN HASSELGREN (EDS) *Reflections on Phenomenography - Toward a Methodology?* Göteborg 1996

110. ELISABETH HESSLEFORS ARKTOFT *I ord och handling. Innebörder av "att anknyta till elevers erfarenheter", uttryckta av lärare.* Göteborg 1996

111. BARBRO STRÖMBERG *Professionellt förhållningssätt hos läkare och sjuksköterskor. En studie av uppfattningar.* Göteborg 1997

112. HARRIET AXELSSON *Våga lära. Om lärare som förändrar sin miljöundervisning.* Göteborg 1997

113. ANN AHLBERG *Children's ways of handling and experiencing numbers.* Göteborg 1997

114. HUGO WIKSTRÖM *Att förstå förändring. Modellbyggande, simulering och gymnasieelevers lärande.* Göteborg 1997

115. DORIS AXELSEN *Listening to recorded music. Habits and motivation among high-school students.* Göteborg 1997.

116. EWA PILHAMMAR ANDERSSON *Handledning av sjuksköterskestuderande i klinisk praktik.* Göteborg 1997

117. OWE STRÅHLMAN *Elitidrott, karriär och avslutning.* Göteborg 1997

118. AINA TULLBERG T*eaching the 'mole'. A phenomenographic inquiry into the didactics of chemistry.* Göteborg 1997.

119. DENNIS BEACH *Symbolic Control and Power Relay Learning in Higher Professional Education.* Göteborg 1997

120. HANS-ÅKE SCHERP *Utmanande eller utmanat ledarskap. Rektor, organisationen och förändrat undervisningsmönster i gymnasieskolan.* Göteborg 1998

121. STAFFAN STUKÁT *Lärares planering under och efter utbildningen.* Göteborg 1998

122. BIRGIT LENDAHLS ROSENDAHL *Examensarbetets innebörder. En studie av blivande lärares utsagor.* Göteborg 1998

123. ANN AHLBERG *Meeting Mathematics. Educational studies with young children.* Göteborg 1998

124. MONICA ROSÉN *Gender Differences in Patterns of Knowledge.* Göteborg 1998.

125. HANS BIRNIK L*ärare- elevrelationen. Ett relationistiskt perspektiv.* Göteborg 1998

126. MARGRETH HILL *Kompetent för "det nya arbetslivet"? Tre gymnasieklasser reflekterar över och diskuterar yrkesförberedande studier.* Göteborg 1998

127. LISBETH ÅBERG-BENGTSSON *Entering a Graphicate Society. Young Children Learning Graphs and Charts.* Göteborg 1998

128. MELVIN FEFFER *The Conflict of Equals: A Constructionist View of Personality Development.* Göteborg 1999

129. ULLA RUNESSON *Variationens pedagogik. Skilda sätt att behandla ett matematiskt innehåll.* Göteborg 1999

130. SILWA CLAESSON *"Hur tänker du då?" Empiriska studier om relationen mellan forskning om elevuppfattningar och lärares undervisning.* Göteborg 1999

131. MONICA HANSEN *Yrkeskulturer i möte. Läraren, fritidspedagogen och samverkan.* Göteborg 1999

132. JAN THELIANDER *Att studera arbetets förändring under kapitalismen. Ure och Taylor i pedagogiskt perspektiv.* Göteborg 1999

133. TOMAS SAAR *Musikens dimensioner - en studie av unga musikers lärande.* Göteborg 1999

134. GLEN HELMSTAD *Understandings of understanding. An inquiry concerning experiential conditions for developmental learning.* Göteborg 1999

135. MARGARETA HOLMEGAARD *Språkmedvetenhet och ordinlärning. Lärare och inlärare reflekterar kring en betydelsefältsövning i svenska som andraspråk.* Göteborg 1999

136. ALYSON MCGEE *Investigating Language Anxiety through Action Inquiry: Developing Good Research Practices.* Göteborg 1999

137. EVA GANNERUD *Genusperspektiv på lärargärning. Om kvinnliga klasslärares liv och arbete.* Göteborg 1999

138. TELLERVO KOPARE *Att rida stormen ut. Förlossningsberättelser i Finnmark och Sápmi.* Göteborg 1999

139. MAJA SÖDERBÄCK *Encountering Parents. Professional Action Styles among Nurses in Pediatric Care.* Göteborg 1999

140. AIRI ROVIO - JOHANSSON *Being Good at Teaching. Exploring different ways of handling the same subject in Higher Education.* Göteborg 1999

141. EVA JOHANSSON *Etik i små barns värld. Om värden och normer bland de yngsta barnen i förskolan.* Göteborg 1999

142. KENNERT ORLENIUS *Förståelsens paradox. Yrkeserfarenhetens betydelse när förskollärare blir grundskollärare.* Göteborg 1999.

143. BJÖRN MÅRDÉN *De nya hälsomissionärerna – rörelser i korsvägen mellan pedagogik och hälsopromotion.* Göteborg 1999

144. MARGARETA CARLÉN *Kunskapslyft eller avbytarbänk? Möten med industriarbetare om utbildning för arbete.* Göteborg 1999

145. MARIA NYSTRÖM *Allvarligt psykiskt störda människors vardagliga tillvaro.* Göteborg 1999

146. ANN-KATRIN JAKOBSSON *Motivation och inlärning ur genusperspektiv. En studie av gymnasieelever på teoretiska linjer/program.* Göteborg 2000

147. JOANNA GIOTA *Adolescents' perceptions of school and reasons for learning.* Göteborg 2000

148. BERIT CARLSTEDT *Cognitive abilities – aspects of structure, process and measurement.* Göteborg 2000

149. MONICA REICHENBERG *Röst och kausalitet i lärobokstexter. En studie av elevers förståelse av olika textversioner.* Göteborg 2000

150. HELENA ÅBERG *Sustainable waste management in households – from international policy to everyday practice. Experiences from two Swedish field studies.* Göteborg 2000

151. BJÖRN SJÖSTRÖM & BRITT JOHANSSON *Ambulanssjukvård. Ambulanssjukvårdares och läkares perspektiv.* Göteborg 2000

152. AGNETA NILSSON *Omvårdnadskompetens inom hemsjukvård – en deskriptiv studie.* Göteborg 2001

153. ULLA LÖFSTEDT *Förskolan som lärandekontext för barns bildskapande.* Göteborg 2001

154. JÖRGEN DIMENÄS *Innehåll och interaktion. Om elevers lärande i naturvetenskaplig undervisning.* Göteborg 2001

155. BRITT MARIE APELGREN *Foreign Language Teachers' Voices. Personal Theories and Experiences of Change in Teaching English as a Foreign Language in Sweden.* Göteborg 2001

156. CHRISTINA CLIFFORDSON *Assessing empathy: Measurement characteristics and interviewer effects.* Göteborg 2001

157. INGER BERGGREN *Identitet, kön och klass. Hur arbetarflickor formar sin identitet.* Göteborg 2001

158. CARINA FURÅKER *Styrning och visioner – sjuksköterskeutbildning i förändring.* Göteborg 2001

159. INGER BERNDTSSON *Förskjutna horisonter. Livsförändring och lärande i samband med synnedsättning eller blindhet.* Göteborg 2001

160. SONJA SHERIDAN *Pedagogical Quality in Preschool. An issue of perspectives.* Göteborg 2001

161. JAN BAHLENBERG *Den otroliga verkligheten sätter spår. Om Carlo Derkerts liv och konstpedagogiska gärning.* Göteborg 2001

162. FRANK BACH *Om ljuset i tillvaron. Ett undervisningsexperiment inom optik.* Göteborg 2001

163. PIA WILLIAMS *Barn lär av varandra. Samlärande i förskola och skola.* Göteborg 2001

164. VIGDIS GRANUM *Studentenes forestillinger om sykepleie som fag og funksjon.* Göteborg 2001

165. MARIT ALVESTAD *Den komplekse planlegginga. Førskolelærarar om pedagogisk planlegging og praksis.* Göteborg 2001

166. GIRMA BERHANU *Learning-In-Context. An Ethnographic Investigation of Mediated Learning Experiences among Ethiopian Jews in Israel.* Göteborg 2001.

167. OLLE ESKILSSON *En longitudinell studie av 10 – 12-åringars förståelse av materiens förändringar.* Göteborg 2001

168. JONAS EMANUELSSON *En fråga om frågor. Hur lärares frågor i klassrummet gör det möjligt att få reda på elevernas sätt att förstå det som undervisningen behandlar i matematik och naturvetenskap.* Göteborg 2001

169. BIRGITTA GEDDA *Den offentliga hemligheten. En studie om sjuksköterskans pedagogiska funktion och kompetens i folkhälsoarbetet.* Göteborg 2001

170. FEBE FRIBERG *Pedagogiska möten mellan patienter och sjuksköterskor på en medicinsk vårdavdelning. Mot en vårddidaktik på livsvärldsgrund.* Göteborg 2001

171. MADELEINE BERGH *Medvetenhet om bemötande. En studie om sjuksköterskans pedagogiska funktion och kompetens i närståendeundervisning.* Göteborg 2002

172. HENRIK ERIKSSON *Den diplomatiska punkten – maskulinitet som kroppsligt identitetsskapande projekt i svensk sjuksköterskeutbildning.* Göteborg 2002

173. SOLVEIG LUNDGREN *I spåren av en bemanningsförändring. En studie av sjuksköterskors arbete på en kirurgisk vårdavdelning.* Göteborg 2002

174. BIRGITTA DAVIDSSON *Mellan soffan och katedern. En studie av hur förskollärare och grundskollärare utvecklar pedagogisk integration mellan förskola och skola.* Göteborg 2002

175. KARI SØNDENÅ *Tradisjon og Transcendens – ein fenomenologisk studie av refleksjon i norsk førskulelærarutdanning.* Göteborg 2002

176. CHRISTINE BENTLEY *The Roots of Variation of English-Teaching. A Phenomenographic Study Founded on an Alternative Basic Assumption.* Göteborg 2002

177. ÅSA MÄKITALO *Categorizing Work: Knowing, Arguing, and Social Dilemmas in Vocational Guidance.* Göteborg 2002

178. MARITA LINDAHL *VÅRDA – VÄGLEDA – LÄRA. Effektstudie av ett interventionsprogram för pedagogers lärande i förskolemiljön.* Göteborg 2002

179. CHRISTINA BERG *Influences on schoolchildren's dietary selection. Focus on fat and fibre at breakfast.* Göteborg 2002

180. MARGARETA ASP *Vila och lärande om vila. En studie på livsvärldsfenomenologisk grund.* Göteborg 2002

181. FERENCE MARTON & PAUL MORRIS (EDS) *What matters? Discovering critical contitions of classroom learning.* Göteborg 2002

182. ROLAND SEVERIN *Dom vet vad dom talar om. En intervjustudie om elevers uppfattningar av begreppen makt och samhällsförändring.* Göteborg 2002

Editors: Björn Andersson, Jan Holmer and Ingrid Pramling Samuelsson

183. MARLÉNE JOHANSSON *Slöjdpraktik i skolan – hand, tanke, kommunikation och andra medierande redskap.* Göteborg 2002

184. INGRID SANDEROTH *Om lust att lära i skolan: En analys av dokument och klass 8y.* Göteborg 2002

185. INGA-LILL JAKOBSSON *Diagnos i skolan. En studie av skolsituationer för elever med syndromdiagnos.* Göteborg 2002

186. EVA-CARIN LINDGREN *Empowering Young Female Athletes – A Possible Challenge to the Male Hegemony in Sport. A Descriptive and Interventional Study.* Göteborg 2002

187. HANS RYSTEDT *Bridging practices. Simulations in education for the health-care professions.* Göteborg 2002

188. MARGARETA EKBORG *Naturvetenskaplig utbildning för hållbar utveckling? En longitudinell studie av hur studenter på grunskollärarprogrammet utvecklar för miljöundervisning relevanta kunskaper i naturkunskap.* Göteborg 2002

189. ANETTE SANDBERG *Vuxnas lekvärld. En studie om vuxnas erfarenheter av lek.* Göteborg 2002

190. GUNLÖG BREDÄNGE *Gränslös pedagog. Fyra studier om utländska lärare i svensk skola.* Göteborg 2003

191. PER-OLOF BENTLEY *Mathematics Teachers and Their Teaching. A Survey Study.* Göteborg 2003

192. KERSTIN NILSSON *MANDAT – MAKT – MANAGEMENT. En studie av hur vårdenhetschefers ledarskap konstrueras.* Göteborg 2003

193. YANG YANG *Measuring Socioeconomic Status and its Effects at Individual and Collective Levels: A Cross-Country Comparison.* Göteborg 2003

194. KNUT VOLDEN *Mediekunnskap som mediekritikk.* Göteborg 2003.

195. LOTTA LAGER-NYQVIST *Att göra det man kan – en longitudinell studie av hur sju lärarstudenter utvecklar sin undervisning och formar sin lärarroll i naturvetenskap.* Göteborg 2003

196. BRITT LINDAHL *Lust att lära naturvetenskap och teknik? En longitudinell studie om vägen till gymnasiet.* Göteborg 2003

197. ANN ZETTERQVIST *Ämnesdidaktisk kompetens i evolutionsbiologi. En intervjuundersökning med no/biologilärare.* Göteborg 2003

198. ELSIE ANDERBERG *Språkanvändningens funktion vid utveckling av kunskap om objekt.* Göteborg 2003.

199. JAN GUSTAFSSON *Integration som text, diskursiv och social praktik. En policyetnografisk fallstudie av mötet mellan skolan och förskoleklassen.* Göteborg 2003.

200. EVELYN HERMANSSON *Akademisering och professionalisering – barnmorskans utbildning i förändring.* Göteborg 2003

201. KERSTIN VON BRÖMSSEN *Tolkningar, förhandlingar och tystnader. Elevers tal om religion i det mångkulturella och postkoloniala rummet.* Göteborg 2003

202. MARIANNE LINDBLAD FRIDH *Från allmänsjuksköterska till specialistsjuksköterska inom intensivvård. En studie av erfarenheter från specialistutbildningen och från den första yrkesverksamma tiden inom intensivvården.* Göteborg 2003

203. BARBRO CARLI *The Making and Breaking of a Female Culture: The History of Swedish Physical Education 'in a Different Voice'.* Göteborg 2003

204. ELISABETH DAHLBORG-LYCKHAGE *"Systers" konstruktion och mumifiering – i TV-serier och i studenters föreställningar.* Göteborg 2003

205. ULLA HELLSTRÖM MUHLI *Att överbrygga perspektiv. En studie av behovsbedömningssamtal inom äldreinriktat socialt arbete.* Göteborg 2003

206. KRISTINA AHLBERG *Synvändor. Universitetsstudenters berättelser om kvalitativa förändringar av sätt att erfara situationers mening under utbildningspraktik.* Göteborg 2004

207. JONAS IVARSSON *Renderings & Reasoning: Studying artifacts in human knowing.* Göteborg 2004

208. MADELEINE LÖWING *Matematikundervisningens konkreta gestaltning. En studie av kommunikationen lärare – elev och matematiklektionens didaktiska ramar.* Göteborg 2004

209. PIJA EKSTRÖM *Makten att definiera. En studie av hur beslutsfattare formulerar villkor för specialpedagogisk verksamhet.* Göteborg 2004

210. CARIN ROOS *Skriftspråkande döva barn. En studie om skriftspråkligt lärande i förskola och skola.* Göteborg 2004

211. JONAS LINDEROTH *Datorspelandets mening. Bortom idén om den interaktiva illusionen.* Göteborg 2004

212. ANITA WALLIN *Evolutionsteorin i klassrummet. På väg mot en ämnesdidaktisk teori för undervisning i biologisk evolution.* Göteborg 2004

213. EVA HJÖRNE *Excluding for inclusion? Negotiating school careers and identities in pupil welfare settings in the Swedish school.* Göteborg 2004

214. MARIE BLIDING *Inneslutandets och uteslutandets praktik. En studie av barns relationsarbete i skolan.* Göteborg 2004

215. LARS-ERIK JONSSON *Appropriating Technologies in Educational Practices. Studies in the Contexts of Compulsory Education, Higher Education, and Fighter Pilot Training.* Göteborg 2004

216. MIA KARLSSON *An ITiS Teacher Team as a Community of Practice.* Göteborg 2004

217. SILWA CLAESSON *Lärares levda kunskap.* Göteborg 2004

218. GUN-BRITT WÄRVIK *Ambitioner att förändra och artefakters verkan. Gränsskapande och stabiliserande praktiker på produktionsgolvet.* Göteborg 2004

219. KARIN LUMSDEN WASS *Vuxenutbildning i omvandling. Kunskapslyftet som ett sätt att organisera förnyelse.* Göteborg 2004

220. LENA DAHL *Amningspraktikens villkor. En intervjustudie av en grupp kvinnors föreställningar på och erfarenheter av amning.* Göteborg 2004

221. ULRIC BJÖRCK *Distributed Problem-Based Learning. Studies of a Pedagogical Model in Practice.* Göteborg 2004

222. ANNEKA KNUTSSON *"To the best of your knowledge and for the good of your neighbour". A study of traditional birth attendants in Addis Ababa, Ethiopia.* Göteborg 2004

223. MARIANNE DOVEMARK *Ansvar – flexibilitet – valfrihet. En etnografisk studie om en skola i förändring.* Göteborg 2004

224. BJÖRN HAGLUND *Traditioner i möte. En kvalitativ studie av fritidspedagogers arbete med samlingar i skolan.* Göteborg 2004

225. ANN-CHARLOTTE MÅRDSJÖ *Lärandets skiftande innebörder – uttryckta av förskollärare i vidareutbildning.* Göteborg 2005

226. INGRID GRUNDÉN *Att återerövra kroppen. En studie av livet efter en ryggmärgsskada.* Göteborg 2005

227. KARIN GUSTAFSSON & ELISABETH MELLGREN *Barns skriftspråkande – att bli en skrivande och läsande person.* Göteborg 2005

228. GUNNAR NILSSON *Att äga π. Praxisnära studier av lärarstudenters arbete med geometrilaborationer.* Göteborg 2005.

229. BENGT LINDGREN *Bild, visualitet och vetande. Diskussion om bild som ett kunskapsfält inom utbildning.* Göteborg 2005

230. PETRA ANGERVALL *Jämställdhetsarbetets pedagogik. Dilemman och paradoxer i arbetet med jämställdhet på ett företag och ett universitet.* Göteborg 2005

231. LENNART MAGNUSSON *Designing a responsive support service for family carers of frail older people using ICT.* Göteborg 2005

232. MONICA REICHENBERG *Gymnasieelever samtalar kring facktexter. En studie av textsamtal med goda och svaga läsare.* Göteborg 2005

233. ULRIKA WOLFF *Characteristics and varieties of poor readers.* Göteborg 2005

234. CECILIA NIELSEN *Mellan fakticitet och projekt. Läs- och skrivsvårigheter och strävan att övervinna dem.* Göteborg 2005.

235. BERITH HEDBERG *Decision Making and Communication in Nursing Practice. Aspects of Nursing Competence.* Göteborg 2005

236. MONICA ROSÉN, EVA MYRBERG & JAN-ERIC GUSTAFSSON *Läskompetens i skolår 3 och 4. Nationell rapport från PIRLS 2001 i Sverige. The IEA Progress in International Reading Literacy Study.* Göteborg 2005

237. INGRID HENNING LOEB *Utveckling och förändring i kommunal vuxenutbildning. En yrkeshistorisk ingång med berättelser om lärarbanor.* Göteborg 2006.

238. NIKLAS PRAMLING *Minding metaphors: Using figurative language in learning to represent.* Göteborg 2006

239. KONSTANTIN KOUGIOUMTZIS *Lärarkulturer och professionskoder. En komparativ studie av idrottslärare i Sverige och Grekland.* Göteborg 2006

240. STEN BÅTH *Kvalifikation och medborgarfostran. En analys av reformtexter avseende gymnasieskolans samhällsuppdrag.* Göteborg 2006.

241. EVA MYRBERG *Fristående skolor i Sverige – Effekter på 9-10-åriga elevers läsförmåga.* Göteborg 2006

242. MARY-ANNE HOLFVE-SABEL *Attitudes towards Swedish comprehensive school. Comparisons over time and between classrooms in grade 6.* Göteborg 2006

243. CAROLINE BERGGREN *Entering Higher Education – Gender and Class Perspectives.* Göteborg 2006

244. CRISTINA THORNELL & CARL OLIVESTAM *Kulturmöte i centralafrikansk kontext med kyrkan som arena.* Göteborg 2006

245. ARVID TREEKREM *Att leda som man lär. En arbetsmiljöpedagogisk studie av toppledares ideologier om ledarskapets taktiska potentialer.* Göteborg 2006

246. EVA GANNERUD & KARIN RÖNNERMAN *Innehåll och innebörd i lärares arbete i förskola och skola – en fallstudie ur ett genusperspektiv.* Göteborg 2006

247. JOHANNES LUNNEBLAD *Förskolan och mångfalden – en etnografisk studie på en förskola i ett multietniskt område.* Göteborg 2006

248. LISA ASP-ONSJÖ *Åtgärdsprogram – dokument eller verktyg? En fallstudie i en kommun.* Göteborg 2006

249. EVA JOHANSSON & INGRID PRAMLING SAMUELSSON *Lek och läroplan. Möten mellan barn och lärare i förskola och skola.* Göteborg 2006

250. INGER BJÖRNELOO *Innebörder av hållbar utveckling. En studie av lärares utsagor om undervisning.* Göteborg 2006

251. EVA JOHANSSON *Etiska överenskommelser i förskolebarns världar.* Göteborg 2006

252. MONICA PETERSSON *Att genuszappa på säker eller osäker mark. Hem- och konsumentkunskap ur ett könsperspektiv.* Göteborg 2007

253. INGELA OLSSON *Handlingskompetens eller inlärd hjälplöshet? Lärandeprocesser hos verkstadsindustriarbetare.* Göteborg 2007

254. HELENA PEDERSEN *The School and the Animal Other. An Ethnography of human-animal relations in education.* Göteborg 2007

255. ELIN ERIKSEN ØDEGAARD *Meningsskaping i barnehagen. Innhold og bruk av barns og voksnes samtalefortellinger.* Göteborg 2007

256. ANNA KLERFELT *Barns multimediala berättande. En länk mellan mediakultur och pedagogisk praktik.* Göteborg 2007

257. PETER ERLANDSON *Docile bodies and imaginary minds: on Schön's reflection-in-action.* Göteborg 2007

258. SONJA SHERIDAN OCH PIA WILLIAMS *Dimensioner av konstruktiv konkurrens. Konstruktiva konkurrensformer i förskola, skola och gymnasium.* Göteborg 2007

259. INGELA ANDREASSON *Elevplanen som text - om identitet, genus, makt och styrning i skolans elevdokumentation.* Göteborg 2007

Editors: Jan-Eric Gustafsson, Annika Härenstam and Ingrid Pramling Samuelsson

260. ANN-SOFIE HOLM *Relationer i skolan. En studie av femininiteter och maskuliniteter i år 9.* Göteborg 2008

261. LARS-ERIK NILSSON *But can't you see they are lying: Student moral positions and ethical practices in the wake of technological change.* Göteborg 2008

262. JOHAN HÄGGSTRÖM *Teaching systems of linear equations in Sweden and China: What is made possible to learn?* Göteborg 2008

263. GUNILLA GRANATH *Milda makter! Utvecklingssamtal och loggböcker som disciplineringstekniker.* Göteborg 2008

264. KARIN GRAHN *Flickor och pojkar i idrottens läromedel. Konstruktioner av genus i ungdomsträningsutbildningen.* Göteborg 2008.

265. PER-OLOF BENTLEY *Mathematics Teachers and Their Conceptual Models. A New Field of Research.* Göteborg 2008

266. SUSANNE GUSTAVSSON *Motstånd och mening. Innebörd i blivande lärares seminariesamtal.* Göteborg 2008

267. ANITA MATTSSON *Flexibel utbildning i praktiken. En fallstudie av pedagogiska processer i en distansutbildning med en öppen design för samarbetslärande.* Göteborg 2008

268. ANETTE EMILSON *Det önskvärda barnet. Fostran uttryckt i vardagliga kommunikationshandlingar mellan lärare och barn i förskolan.* Göteborg 2008

269. ALLI KLAPP LEKHOLM *Grades and grade assignment: effects of student and school charachterisitcs.* Göteborg 2008

270. ELISABETH BJÖRKLUND *Att erövra litteracitet. Små barns kommunikativa möten med berättande, bilder, text och tecken i förskolan.* Göteborg 2008

271. EVA NYBERG *Om livets kontinuitet. Undervisning och lärande om växters och djurs livscykler - en fallstudie i årskurs 5.* Göteborg 2008

272. CANCELLED

273. ANITA NORLUND *Kritisk sakprosaläsning i gymnasieskolan. Didaktiska perspektiv på läroböcker, lärare och nationella prov.* Göteborg 2009

274. AGNETA SIMEONSDOTTER SVENSSON *Den pedagogiska samlingen i förskoleklasen. Barns olika sätt att erfara och hantera svårigheter.* Göteborg 2009

275. ANITA ERIKSSON *Om teori och praktik i lärarutbildningen. En etnografisk och diskursanalytisk studie.* Göteborg 2009

276. MARIA HJALMARSSON *Lärarprofessionens genusordning. En studie av lärares uppfattningar om arbetsuppgifter, kompetens och förväntningar.* Göteborg 2009.

277. ANNE DRAGEMARK OSCARSON *Self-Assessement of Writing in Learning English as a Foreign Language. A Study at the Upper Secondary School Level.* Göteborg 2009

278. ANNIKA LANTZ-ANDERSSON *Framing in Educational Practices. Learning Activity, Digital Technology and the Logic of Situated Action.* Göteborg 2009

279. RAUNI KARLSSON *Demokratiska värden i förskolebarns vardag.* Göteborg 2009

280. ELISABETH FRANK *Läsförmågan bland 9-10-åringar. Betydelsen av skolklimat, hem- och skolsamverkan, lärarkompetens och elevers hembakgrund.* Göteborg 2009

281. MONICA JOHANSSON *Anpassning och motstånd. En etnografisk studie av gymnasieelevers institutionella identitetsskapande.* Göteborg 2009

282. MONA NILSEN *Food for Thought. Communication and the transformation of work experience in web-based in-service training.* Göteborg 2009

283. INGA WERNERSSON (RED) *Genus i förskola och skola. Förändringar i policy, perspektiv och praktik.* Göteborg 2009

284. SONJA SHERIDAN, INGRID PRAMLING SAMUELSSON & EVA JOHANSSON (RED) *Barns tidiga lärande. En tvärsnittsstudie om förskolan som miljö för barns lärande.* Göteborg 2009

285. MARIE HJALMARSSON *Lojalitet och motstånd - anställdas agerande i ett föränderligt hemtjänstarbete.* Göteborg 2009.

286. ANETTE OLIN *Skolans mötespraktik - en studie om skolutveckling genom yrkesverksammas förståelse.* Göteborg 2009

287. MIRELLA FORSBERG AHLCRONA *Handdockans kommunikativa potential som medierande redskap i förskolan.* Göteborg 2009

288. CLAS OLANDER *Towards an interlanguage of biological evolution: Exploring students´ talk and writing as* an arena for sense-making. Göteborg 2010

Editors: Jan-Eric Gustafsson, Åke Ingerman and Ingrid Pramling Samuelsson

289. PETER HASSELSKOG *Slöjdlärares förhållningssätt i undervisningen.* Göteborg 2010

290. HILLEVI PRELL *Promoting dietary change. Intervening in school and recognizing health messages in commercials.* Göteborg 2010

291. DAVOUD MASOUMI *Quality Within E-learning in a Cultural Context. The case of Iran.* Göteborg 2010

292. YLVA ODENBRING *Kramar, kategoriseringar och hjälpfröknar. Könskonstruktioner i interaktion i förskola, förskoleklass och skolår ett.* Göteborg 2010

293. ANGELIKA KULLBERG *What is taught and what is learned. Professional insights gained and shared by teachers of mathematics.* Göteborg 2010

294. TORGEIR ALVESTAD *Barnehagens relasjonelle verden - små barn som kompetente aktører i produktive forhandlinger.* Göteborg 2010

295. SYLVI VIGMO *New spaces for Language Learning. A study of student interaction in media production in English.* Göteborg 2010

296. CAROLINE RUNESDOTTER *I otakt med tiden? Folkhögskolorna i ett föränderligt fält.* Göteborg 2010

297. BIRGITTA KULLBERG *En etnografisk studie i en thailändsk grundskola på en ö i södra Thailand. I sökandet efter en framtid då nuet har nog av sitt.* Göteborg 2010

298. GUSTAV LYMER *The work of critique in architectural education.* Göteborg 2010

299. ANETTE HELLMAN *Kan Batman vara rosa? Förhandlingar om pojkighet och normalitet på en förskola.* Göteborg 2010

300. ANNIKA BERGVIKEN-RENSFELDT *Opening higher education. Discursive transformations of distance and higher education government.* Göteborg 2010

301. GETAHUN YACOB ABRAHAM *Education for Democracy? Life Orientation: Lessons on Leadership Qualities and Voting in South African Comprehensive Schools.* Göteborg 2010

302. LENA SJÖBERG *Bäst i klassen? Lärare och elever i svenska och europeiska policytexter.* Göteborg 2011

303. ANNA POST *Nordic stakeholders and sustainable catering.* Göteborg 2011

304. CECILIA KILHAMN *Making Sense of Negative Numbers.* Göteborg 2011

305. ALLAN SVENSSON (RED) *Utvärdering Genom Uppföljning. Longitudinell individforskning under ett halvsekel.* Göteborg 2011

306. NADJA CARLSSON *I kamp med skriftspråket. Vuxenstuderande med läs- och skrivsvårigheter i ett livsvärldsperspektiv.* Göteborg 2011

307. AUD TORILL MELAND *Ansvar for egen læring. Intensjoner og realiteter ved en norsk videregående skole.* Göteborg 2011

308. EVA NYBERG *Folkbildning för demokrati. Colombianska kvinnors perspektiv på kunskap som förändringskraft.* Göteborg 2011

309. SUSANNE THULIN *Lärares tal och barns nyfikenhet. Kommunikation om naturvetenskapliga innehåll i förskolan.* Göteborg 2011

310. LENA FRIDLUND *Interkulturell undervisning – ett pedagogiskt dilemma. Talet om undervisning i svenska som andraspråk och i förberedelseklass.* Göteborg 2011

311. TARJA ALATALO *Skicklig läs- och skrivundervisning i åk 1-3. Om lärares möjligheter och hinder.* Göteborg 2011

312. LISE-LOTTE BJERVÅS *Samtal om barn och pedagogisk dokumentation som bedömningspraktik i förskolan. En diskursanalys.* Göteborg 2011

313. ÅSE HANSSON *Ansvar för matematiklärande. Effekter av undervisningsansvar i det flerspråkiga klassrummet.* Göteborg 2011

314. MARIA REIS *Att ordna, från ordning till ordning. Yngre förskolebarns matematiserande.* Göteborg 2011

315. BENIAMIN KNUTSSON *Curriculum in the Era of Global Development – Historical Legacies and Contemporary Approaches.* Göteborg 2011

316. EVA WEST *Undervisning och lärande i naturvetenskap. Elevers lärande i relation till en forskningsbaserad undervisning om ljud, hörsel och hälsa.* Göteborg 2011

317. SIGNILD RISENFORS *Gymnasieungdomars livstolkande.* Göteborg 2011

318. EVA JOHANSSON & DONNA BERTHELSEN (Ed.) *Spaces for Solidarity and Individualism in Educational Contexts.* Göteborg 2012

319. ALASTAIR HENRY *L3 Motivation.* Göteborg 2012

320. ANN PARINDER *Ungdomars matval – erfarenheter, visioner och miljöargument i eget hushåll.* Göteborg 2012

321. ANNE KULTTI *Flerspråkiga barn i förskolan: Villkor för deltagande och lärande.* Göteborg 2012

322. BO-LENNART EKSTRÖM *Kontroversen om DAMP. En kontroversstudie av vetenskapligt gränsarbete och översättning mellan olika kunskapsparadigm.* Göteborg 2012

323. MUN LING LO *Variation Theory and the Improvement of Teaching and Learning.* Göteborg 2012

324. ULLA ANDRÉN *Self-awareness and self-knowledge in professions. Something we are or a skill we learn.* Göteborg 2012

325. KERSTIN SIGNERT *Variation och invarians i Maria Montessoris sinnesträning materiel.* Göteborg 2012

326. INGEMAR GERRBO *Idén om en skola för alla och specialpedagogisk organisering i praktiken.* Göteborg 2012

327. PATRIK LILJA *Contextualizing inquiry. Negotiations of tasks, tools and actions in an upper secondary classroom.* Göteborg 2012

328. STEFAN JOHANSSON *On the Validity of Reading Assessments: Relationships Between Teacher Judgements, External Tests and Pupil Self-assessments.* Göteborg 2013

329. STEFAN PETTERSSON *Nutrition in Olympic Combat Sports. Elite athletes' dietary intake, hydration status and experiences of weight regulation.* Göteborg 2013

330. LINDA BRADLEY *Language learning and technology – student activities in web-based environments.* Göteborg 2013

331. KALLE JONASSON *Sport Has Never Been Modern.* Göteborg 2013

332. MONICA HARALDSSON STRÄNG *Yngre elevers lärande om natur. En studie av kommunikation om modeller i institutionella kontexter.* Göteborg 2013

333. ANN VALENTIN KVIST *Immigrant Groups and Cognitive Tests – Validity Issues in Relation to Vocational Training.* Göteborg 2013

334. ULRIKA BENNERSTEDT *Knowledge at play. Studies of games as members' matters.* Göteborg 2013

335. EVA ÄRLEMALM-HAGSÉR *Engagerade i världens bästa? Lärande för hållbarhet i förskolan.* Göteborg 2013

336. ANNA-KARIN WYNDHAMN *Tänka fritt, tänka rätt. En studie om värdeöverföring och kritiskt tänkande i gymnasieskolans undervisning.* Göteborg 2013

337. LENA TYRÈN *"Vi får ju inte riktigt förutsättningarna för att genomföra det som vi vill." En studie om lärares möjligheter och hinder till förändring och förbättring i praktiken.* Göteborg 2013

Editors: Jan-Eric Gustafsson, Åke Ingerman and Pia Williams

338. ANNIKA LILJA *Förtroendefulla relationer mellan lärare och elev.* Göteborg 2013

339. MAGNUS LEVINSSON *Evidens och existens. Evidensbaserad undervisning i ljuset av lärares erfarenheter.* Göteborg 2013

340. ANNELI SCHWARTZ *Pedagogik, plats och prestationer. En etnografisk studie om en skola i förorten.* Göteborg 2013

341. ELISABET ÖHRN och LISBETH LUNDAHL (red) *Kön och karriär i akademin. En studie inom det utbildningsvetenskapliga fältet.* Göteborg 2013

342. RICHARD BALDWIN *Changing practice by reform. The recontextualisation of the Bologna process in teacher education.* Göteborg 2013

343. AGNETA JONSSON *Att skapa läroplan för de yngsta barnen i förskolan. Barns perspektiv och nuets didaktik.* Göteborg 2013

344. MARIA MAGNUSSON *Skylta med kunskap. En studie av hur barn urskiljer grafiska symboler i hem och förskola.* Göteborg 2013

345. ANNA-LENA LILLIESTAM *Aktör och struktur i historieundervisning. Om utveckling av elevers historiska resonerande.* Göteborg 2013

346. KRISTOFFER LARSSON *Kritiskt tänkande i grundskolans samhällskunskap. En fenomenografisk studie om manifesterat kritiskt tänkande i samhällskunskap hos elever i årskurs 9.* Göteborg 2013

347. INGA WERNERSSON och INGEMAR GERRBO (red) *Differentieringens janusansikte. En antologi från Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet.* Göteborg 2013

348. LILL LANGELOTZ *Vad gör en skicklig lärare? En studie om kollegial handledning som utvecklingspraktik.* Göteborg 2014

349. STEINGERDUR OLAFSDOTTIR *Television and food in the lives of young children.* Göteborg 2014

350. ANNA-CARIN RAMSTEN *Kunskaper som byggde folkhemmet. En fallstudie av förutsättningar för lärande vid teknikskiften inom processindustrin.* Göteborg 2014

351. ANNA-CARIN BREDMAR *Lärares arbetsglädje. Betydelsen av emotionell närvaro i det pedagogiska arbetet.* Göteborg 2014

352. ZAHRA BAYATI *"den Andre" i lärarutbildningen. En studie om den rasifierade svenska studentens villkor i globaliseringens tid.* Göteborg 2014

353 ANDERS EKLÖF *Project work, independence and critical thinking.* Göteborg 2014

354 EVA WENNÅS BRANTE *Möte med multimodalt material. Vilken roll spelar dyslexi för uppfattandet av text och bild?* Göteborg 2014

355 MAGNUS FERRY *Idrottsprofilerad utbildning – i spåren av en avreglerad skola.* Göteborg 2014

356 CECILIA THORSEN *Dimensionality and Predictive validity of school grades: The relative influence of cognitive and socialbehavioral aspects.* Göteborg 2014

Dimensionality and Predictive validity of school grades: The relative influence of cognitive and social-behavioral aspects

In Sweden grades are used in processes of selection to the next educational level. The types of selection instrument used differ, both in different educational systems, and with respect to whether capacity for schooling or previous achievement is preferred. Nevertheless, the predictive validity of school grades, designed to measure previous achievement, has been demonstrated in a multitude of studies. Grades have also been found to predict other outcomes, such as job-performance. Although the reasons why grades display this pattern of predictive power are not fully understood, it is a reasonable assumption that, in part, this can be explained by the influence of both cognitive and social-behavioral aspects. Thus the aim of the present thesis is to investigate the influence of cognitive and social-behavioral aspects on compulsory school grades and the relative importance these aspects have for the predictive power of grades.

The results indicate that both criterion-referenced and norm-referenced compulsory school grades are multidimensional, reflecting both knowledge and skills, and social-behavioral aspects. Dimensions related to knowledge and skills, and a dimension which was common to all grades and all teachers (interpreted in part to reflect social-behavioral aspects) were identified in both grading systems. The multidimensionality of grades was also found to be stable across several birth cohorts. Further, the results suggest that the influence of cognitive abilities on the development of knowledge and skills was substantial, and that there was a continuous influence of fluid abilities throughout compulsory school. All in all, the results indicate that a partial explanation for the predictive power of school grades can be found in the investment of both cognitive and social-behavioral aspects into the acquisition of knowledge and skills, but that there is also a direct influence of social-behavioral aspects on grades as a consequence of teachers' grading.

Cecilia Thorsen has previously worked as a teacher in upper secondary school. She is currently working at the Department of Social and Behavioural Studies at University West, Trollhättan, where she lectures on assessment on a range of teacher education programmes. Her research interests lie within the field of educational assessment and cognitive abilities.

ISBN 978-91-7346-797-1 (print)
ISBN 978-91-7346-798-8 (pdf)
ISSN 0436-1121